FORECASTING RIDERSHIP IMPACTS OF TRANSIT ORIENTED DEVELOPMENT AT MARTA RAIL STATIONS

A Thesis
Presented to
The Academic Faculty

By

George Maier

In Partial Fulfillment
Of the Requirements for the Degree
Master of City and Regional Planning

Georgia Institute of Technology

December, 2015

FORECASTING RIDERSHIP IMPACTS OF TRANSIT ORIENTED DEVELOPMENT AT MARTA RAIL STATIONS

Approved By:

Dr. Randall L. Guensler, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Timothy F. Welch, Advisor
School of City and Regional Planning
Georgia Institute of Technology

Dr. Kari Edison Watkins
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: December 2, 2015

ACKNOWLEDGEMENTS

It has been a privilege to work with a team of professors who without, this thesis would not have been possible. I am grateful to have worked with each of them who have added unique professional perspectives to this research. Specifically, I would like to thank Dr. Welch for the ceaseless support and discerning approach to problem solving. I would like to thank Dr. Guensler for encouraging the relentless pursuit of discovery. I would like to thank Dr. Watkins for her infectious enthusiasm and keen insight of the subject material.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	American Community Survey
ANOVA	Analysis of Variance
ARC	Atlanta Regional Commission
BRP	Binary recursive partitioning
BRT	Bus Rapid Transit
CART	Classification and Regression Trees
CATS	Cherokee Area Transportation System
CBD	Central Business District
CCT	Cobb Community Transit
CTRAN	Clayton Transit
DRM	Direct Ridership Model
ESRI	Environmental Systems Research Institute
GCT	Gwinnett County Transit
GIS	Geographic Information System
GPS	Global Positioning System
GLUP	General Land Use Plan
GRTA	Georgia Regional Transportation Authority
HAT	Hall Area Transit
HLM	Hierrarchical Linear Model
LUMI	Land-Use Mix Index
MARTA	Metropolitan Atlanta Rapid Transit Authority

MPO	Metropolitan Planning Organization
NAICS	North American Industry Classification System
OLS	Ordinary Least Squares
SPSS	Statistical Package for the Social Sciences
TAZ	Traffic Analysis Zone
TCQSM	Transit Capacity and Quality of Service
TDM	Travel Demand Model
TIGER	Topologically Integrated Geographic Encoding and Referencing
TOD	Transit Oriented Development
WMATA	Washington Metropolitan Area Transit Authority

SUMMARY

The Metropolitan Atlanta Rapid Transit Authority (MARTA) Transit Oriented Development (TOD) program has been expanding the number of stations being considered for development of surface parking lots and into the air rights over certain rail stations. As of 2015, MARTA has six rail stations in various stages of TOD development, which will increase multi-modal options for metro Atlanta residents. The overarching goal of TOD development is to increase transit ridership and reduce auto-dependency; hence quantifying the potential benefits of TOD development in terms of ridership is paramount. Despite several drawbacks, travel demand models have historically been utilized to forecast ridership for land use changes and transit improvements. Direct ridership models (DRMs) are transit demand forecasting methods that can be applied to land development in cases where traditional travel demand models (TDMs) are not well suited.

DRMs leverage geographic tools commonly used by planners to take advantage of small scale pedestrian environment factors immediately surrounding transit stations. Although DRM data and methods can achieve greater precision in predicting local walk-access transit trips, the lack of regional and large-scale datasets reduces the ability to model ridership generated from riders outside the immediate vicinity of the rail stations. Stations that have high multi-modal access trips, particularly via personal vehicle and connecting buses, are not typically accounted for by DRMs. Hence, this study focuses on pedestrian-based rail boardings only, a metric that also allows the use of a large scale

onboard survey distributed by the Atlanta Regional Commission (ARC) in late 2009 and early 2010 in Atlanta, Georgia.

Analysis of the large scale on-board ridership survey also reveals variables that may be useful in forecasting ridership at the station level when coupled with available census data. Comparison of variables such as income, age, gender, ethnicity, and race from census data with the large scale survey guided the selection of candidate variables to be included in a DRM for MARTA rail stations. Results from the comparison showed that using census data in DRMs does not always accurately reflect the ridership demographics. Notable differences in pedestrian-based ridership and transit catchments appear to occur in populations making less than $40,000, African American populations, and the young and elderly populations. Large differences in the survey and census data reported around the stations raise questions about the usability of census data in predicting ridership at rail stations.

Despite the shortcomings of using census data to directly predict walk access transit ridership, an ordinary least squared (OLS) regression model predicts a high proportion of variance of pedestrian-based ridership in Atlanta, Georgia. A small number of variables were incorporated into a DRM to show the strong relationship of employment density with pedestrian based ridership. The number of low income residents was also influential in increasing ridership via walk access.

CHAPTER 1

INTRODUCTION

Transit oriented development (TOD) has been an effective development strategy for transit agencies to maintain steady ridership levels within transit systems and increase local transit demand. Increasing the density of development around transit stations has been a preferred growth alternative in many cities over the past several decades, with new emphasis placed on this type of development as a fundamental aspect of the smart growth movement in the U.S., such as Arlington, San Francisco, and Salt Lake City. In the 1960's, the planning of five Washington Metropolitan Area Transit Authority (WMATA) rail stations in Arlington, Virginia inspired changes to the county's General Land Use Plan (GLUP) to develop TOD. Instead of aligning the rail stations in the median of Interstate 66, planners and officials opted to locate the stations in town centers. Planners created a "Bulls Eye Concept", shown in Figure 1 below, where the land use plan called for the densest development to occur within 0.25 miles from the rail station (Brosnan, 2010). Now known as the Rosslyn-Ballston corridor, the growth strategy is hailed as a great success. From 1970 to 2009, employment around stations increased by 76,500 jobs, office space increased by 16.2 million square feet, and the number of housing units increased by 21,643. From 1991 to 2008, the ridership at each of the station more than doubled (Brosnan, 2010).

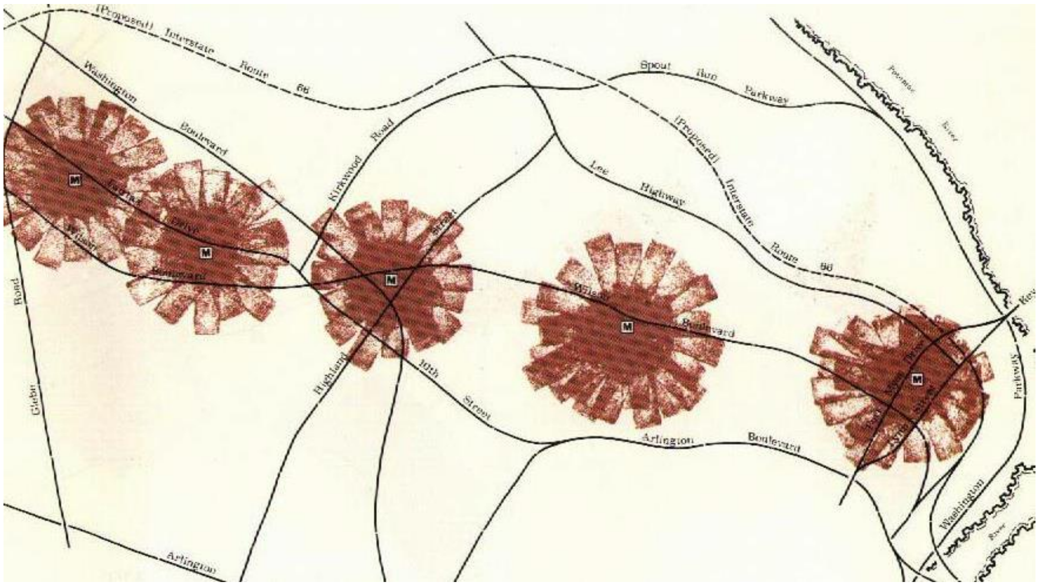

Figure 1: TOD bull’s eye concept in Arlington, Virginia (Brosnan, 2010).

The adoption of TOD strategies in other cities however, has been slow until recently. New emphasis has been placed on densifying development around transit stations as a fundamental aspect of the smart growth movement in the U.S. In mid-2005, the Bay Area Rapid Transit (BART) adopted a TOD policy outlining specific goals to increase ridership, enhance the quality of life, grow its financial base, and reduce automobile access mode share by using development near and in the transit stations. Since then, BART has completed a total of eight TODs, including 867 housing units, 72,600 square feet of commercial property, and 107,000 square feet of office space (Goll, 2015). BART officials estimate that the development has resulted in an additional 379,000 annual trips and $1.4 million in revenue (Goll, 2015).

In Atlanta, the Metropolitan Atlanta Rapid Transit Authority (MARTA) has recently gained interest in the potential of TOD around rail stations. The Office of TOD and Real Estate has increased its efforts to develop land on, around and even above

MARTA stations in hopes of capitalizing on this movement. In an update letter of the multiple TOD initiates, Senior Director of transit oriented development Amanda Rhein states that general manager/chief executive officer Keith Parker has a vision of "growing the agency's TOD program as a way of building our building ridership, increasing revenues, and enhancing communities" (Rhein, 2015).

Quantifying the benefits of TOD in terms of ridership has traditionally been assessed using four-step travel demand models (Cervero et al., 2010). Although regional four-step models excel in many aspects of transportation forecasting, there are several well-known issues with predicting transit ridership with these models. The primary focus of regional models is to show the flow of major trip producers and trip attractors between zones across a metro region by different modes of transport, the vast majority being by personal automobile. Because these models are regional in scope they require a large amount of data and computing power. To ease the computational burden the units are typically large and do not incorporate pedestrian level detail. Therefore the fundamental elements of TOD such as increased pedestrian accessibility and high density development within a short walking distance to the station are not emphasized in forecasting transit ridership within regional four-step models.

One response to the challenges of modeling ridership at the transit station level has been the use of direct ridership models (DRMs), also known as direct estimation or direct demand models. These models are typically much simpler, less expensive, and quicker to run and modify than four-step models largely because they utilize station level characteristics to forecast ridership. The incorporation of station level details, such as adjacent land use and walking access, give DRMs a significant advantage in evaluating the potential effects of TOD development on local transit ridership. Although there are several drawbacks to these models, such as neglecting regional movement and mode comparison, the research on DRMs has shown that they can be effective for sketch planning purposes (Cervero, 2006). While it is not recommended that DRMs replace regional models for transit ridership forecasting, they are emerging as useful tools for supplementing the four-step model approach (Cervero et al., 2010). This thesis applies the DRM approach to MARTA station data, the majority of which are freely available via the census, and demonstrates the ability of such data to predict pedestrian based ridership.

1.1 Outline

Background information on traditional four-step modeling and DRMs is discussed in Chapter 2 of this thesis, including the advantages and disadvantages of each approach in the context of transit ridership forecasting. DRMs are then examined in further detail, revealing the common components from a review of existing literature. Chapter 2 also reviews the various methodologies for defining and measuring transit catchment areas and their effect on transit accessibility metrics. Chapter 3 defines the data and methodologies used in creating a DRM for MARTA rail stations, with results detailed in Chapter 4. Chapter 5 discusses conclusions of this study and speaks to the broader implications of this work on transit ridership forecasting and transit planning practices such as TOD.

CHAPTER 2

BACKGROUND

Over the past several years there has been a surge of interest in the potential development on and around Metropolitan Atlanta Regional Transit Authority (MARTA) rail stations. Although MARTA's first attempt at transit oriented development (TOD), Lindbergh Center Station, has received mixed reviews, support for TOD is growing in the metro area. In 2012, MARTA's Office of TOD and Real Estate targeted a total of ten rail stations as potential sites for mixed-use developments. As of 2015, six rail stations are in various stages of the development timeline. Five of the rail stations (Avondale, Chamblee, Brookhaven/Oglethorpe, Edgewood/Candler, and King Memorial) have development partners, while Oakland City station is in the proposal process. The vision for these developments incorporates building on surface parking lots with multi-story office, residential, and retail and replacing the parking needs with parking decks. MARTA initiated TOD opportunities in 2014 by releasing a Request for Expression of Interest for air rights to develop directly above four more stations: Arts Center, Lenox, Midtown, and North Avenue (Rhein, 2015).

In anticipation of the development of transit stations into centers of activity, ridership benefits are typically quantified using regional four-step travel demand models (TDM). There are, however, several caveats when using traditional four-step modeling to predict transit ridership. Compared to using a DRM, the process of modeling ridership through regional four-step models is expensive both in terms of time and money. Direct ridership models (DRMs) show particular promise in modeling station environments and other small scale factors that can affect ridership. More specifically, DRMs have been evolving to include more advanced techniques that may lend some valuable insight into the relationship of transit ridership and the immediately adjacent built environment. While regional travel demand models have certain major advantages in modeling ridership, DRMs may be better suited in modeling the benefits of TOD. An explanation the four-step model approach is below, followed by detailed commentary on DRMs including the advantages and disadvantages of each method.

2.1 Traditional four-step models

Transportation infrastructure investment requires analytical procedures to predict outcomes of various alternatives. In the case of planning major highways and regional connections with roadways, the most commonly used approach has been the traditional four-step TDM. Metropolitan planning organizations (MPOs) are required to maintain and update the TDMs for these investments (Meyer & Miller, 2001). The focus of the four-step model has been (and still is) the regional movement of vehicles throughout the road network, and therefore is constructed to handle large amounts of regional traffic to and from distinct areas called traffic analysis zones (TAZ) (Cervero, 2006). The TAZs are designed to incorporate similar land uses within TAZs, and separate dissimilar land uses across TAZs, and typically use roads as boundaries, resulting in a wide range of shapes and sizes. Population and employment data are attributed to centroids in each TAZ to generate travel demand, which informs produced and attracted trips using a gravity model. Each trip is then connected to a TAZ centroid from a simplified highway or transit network link. While the four-step TDM may be an effective method for conducting mesoscale or macro-scale travel movement within a region, the coarse scale of TAZs limit its ability to model fine grained details of land use characteristics (Cervero, 2006). A prime example of this is the use of four-step models to forecast ridership at transit stations without detailed information on station characteristics.

Gutierrez et al. (2011) argue that the use of TDM models for transit ridership forecasting has several potential issues, the first of which is model accuracy. This concern about accuracy arises from the greater attention paid to matching traffic counts on roadway segments than on matching line loads on transit routes or stations, potentially leading to more accurate traffic counts at the expense of transit ridership counts. Another concern voiced by Gutierrez et al. (2011) is on the relatively small number of transit trips that serve as travel input data from household surveys, leading to greater uncertainly for transit ridership. Land-use sensitivity is also a concern, because many regional models are not responsive to land-use changes. Lastly, the costs associated with building and running a full scale four-step model can be prohibitively expensive in terms of time and money (Gutierrez et al., 2011).

Comparison of model skims with observed or likely path data may reveal the shortcomings of using TDMs to predict transit level of service. Work by Zuehlke (2007) in Atlanta compared model-reported skims from a four-step TDM (ARC 20-county base year model) to global positioning system (GPS)-revealed automobile skims and transit traces. The results of the transit traces comparison showed that the transit traces were about 24% longer than the minimum modeled transit network skims, indicating that the actual transit path for a passenger is oftentimes much longer than what the TDM predicts. The methods for assessing walk access are also over simplified, which is coded in the TDMs as a walk-to-transit impedance value. Zones are split into market segments, which are coded as either walk or non-walk based on the straight line distance of 0.125 mile by 0.125 mile grid cells. For the Atlanta Regional Model (ARC) model, grid cells are classified as accessible to transit if they are within 0.4 straight-line miles away from a transit stop. Zuehlke (2007) notes that this methodology may be reasonable for use in a regional travel demand model but has shortcomings as well. Walk impedances from TAZ centroids to transit nodes do not account for the variability in real or perceived walk path experience. Moreover, because of the zonal representation, the TDM cannot model the heterogeneous nature of land use and trip generation (Zuehlke, 2007).

2.2 DRMs

The challenges associated with forecasting transit ridership via the traditional four-step modeling approach have motivated researchers and transit professionals to investigate complimentary transit ridership forecasting methodologies. One such approach that addresses many of the aforementioned concerns is direct ridership modeling using multiple regression analysis. Typically this approach involves using station environment, services features and other directly quantifiable characteristics of transit stations to estimate ridership generated at each location (Kuby et al., 2004; Chu, 2004; Cervero, 2006; Gutierrez et al., 2011).

DRMs are constructed using station level components as independent variables in model construction and transit ridership as the dependent variable, also at the station level. Compilation of the results from several stations can yield results for an entire corridor (Gutierrez et al., 2011). Using multiple regression for the model enables researchers to construct DRMs with a multitude of variables, with the flexibility to supplement or remove variables with relative ease. Moreover, the results are understood by a wider audience because of the widespread application of regression models (Kuby et al., 2004). The ease of interpretation removes much of the abstruseness of travel demand models that public policy and decision-makers experience.

Gutierrez et al. (2011) states that the station level components incorporated into DRMs fall into one of three categories: the built environment, socioeconomic, or station characteristics. The built environment element includes the density of the area, because more people living and working within walking distance to a station leads to a greater likelihood of station patronage (Murray et al., 1998; Cervero, 2002). Another key aspect of the built environment that affects transit ridership is the land-use mix and type. Research by Filion (2001) showed that mixed-use suburban development encourages greater ridership than typical suburban development. A variety of land uses around the station can also increase the likelihood of a lower peak to base ratio of transit ridership, meaning transit ridership will be more balanced throughout the day rather than used primarily only during peak travel times (Cervero, 2006). Accessibility is strongly affected by the built environment, and therefore is another important component in determining ridership. Shorter blocks and frequent streets can improve the accessibility to transit stations and therefore arguably ridership (Hsiao et al., 1997; Loutzenheiser, 1997).

In addition to built environment factors, research shows that certain socioeconomic variables affect transit ridership as well. Household income and car ownership are two such variables that are negatively associated with transit ridership, meaning those with vehicles available or more income are less likely to take transit (Chow et al., 2006). Places with high concentration of racial and ethnic minorities and immigrants on the other hand, are positively associated with transit ridership rates (Jin, 2005). Factors like gender and age can also affect an individual's likelihood of taking transit in certain conditions (Chu, 2004).

Station characteristics comprise another group of factors that are commonly utilized in DRM multiple regression (Gutierrez et al., 2011). The relation of the station to the rest of the system (i.e. if the station type is intermediate, terminal, interchange, or intermodal), affects ridership. Terminal stations, stations at the end of the transit line, typically receive an increase in ridership because of the lack of other stations encroaching on catchment area (O'Sullivan and Morrall, 1996). Interchange and intermodal stations also typically receive a ridership bonus because of the connections associated with each type of station (Gutierrez et al., 2011). Likewise, park-and-ride facilities located at stations may increase ridership (Kuby et al., 2004). Station spacing that is greater in distance increases the geographic area that is allocated to each station, and therefore may contribute to increased ridership per station (Kuby et al., 2004). The effect of distance from the station to the central business district and the center of the network has also been observed, showing that closer stations have greater ridership than more distant ones (Kuby et al., 2004). Lastly, service frequency is linked to increased ridership. However, Taylor and Fink (2003) question the use of this variable as an indicator of demand because increased service frequency is often scheduled to meet demand. The utility of DRMs may be limited in deciphering the relationship of increased demand resulting from increased service supply.

2.2.1 Advantages and disadvantages of DRMs

The simplicity of DRMs afford many advantages over four-step models, including quicker and more flexible ridership modeling. Because DRMs are constructed using multiple regression analysis, they can be run quickly with personal computers and then modified and rerun with relatively little computing power and time. This provides planners and researchers the flexibility to test multiple alternatives quickly and economically. Testing several scenarios becomes viable with this approach, affording more options in attempting to discover influential factors affecting ridership. Kuby et al. (2004), for example, states that direct demand models are useful for experimenting with alternative alignments of light rail transit.

DRMs incorporate detailed datasets of surrounding land-use characteristics and therefore are better suited to assess impacts of the built environment on transit ridership. To define the area surrounding the station that is included in the models, transit catchments are delineated using various geographic information systems (GIS) techniques that show the service area around each station. This enables DRMs to use pedestrian-scale measurements and TOD variables to predict ridership outcomes among alternatives. Regional scale TDMs lack this granularity. Instead, TDMs code accessibility of TAZs, assigning productions and attractions based on population and employment assigned to each TAZ. Hence, DRMs are an attractive choice for evaluating the effects of station catchment characteristics on ridership. While DRMs do not benefit from the complexity or scope of four-step models, the flexibility provides a myriad of options to experiment with new variables.

Walters and Cervero (2003) describe the utility of direct demand models as a complementary tool to four-step models, addressing many of the areas that four-step models do not: (1) DRMs are able to measure the effects of combinations of alignments, station locations, and vehicle technology types, (2) DRMs can quickly evaluate variations in parking, feeder bus service, station spacing and transit speed and frequency, (3) DRMs are able to capture the effects of local land-use characteristics like density and walkability of transit catchment areas, (4) DRMs can measure ridership share of competing transit services through corridors, and (5) the timeframe of development to implementation is relatively quick (Walter and Cervero, 2003).

There are however, several disadvantages in using DRMs compared to traditional four-step modeling (Cervero et al., 2010). Comparative travel times of different modes of travel are not included in DRMs, which is a major factor in the mode choice step in traditional regional models. Another important variable in traditional models that is neglected in DRMs is the comparative costs of different modes of travel in terms of time and price. Comparisons to transit's greatest competitor, the private automobile, in terms of accessibility to jobs and shopping is absent as well. Because each observation is a station or stop, the sample sizes used in DRM studies are typically small and therefore the degree of freedom constraints limit the number of variables that can be included in DRMs. The transferability of a DRM from one city to another is not advisable, as the strength and applicability of variables from one region may not be valid in another (Cervero et al., 2010). This limitation forces planners and researchers to recreate DRMs for each region which may or may not include variables that perform well in others.

2.2.2 DRM literature

While DRMs show promise as ridership forecasting tools, the application in peer reviewed studies is still somewhat limited. One early study however, completed by Parsons Brinkhoff (1996), used data from thirteen light-rail systems from the U.S. and Canada including 261 stations. Independent variables like population and employment density within 0.5 miles from the stations, catchment size, distance to CBD (Central Business District), and dummy variables for station characteristics were used in the cross-sectional regression model (Parsons Brinkerhoff, 1996). All but park-and-ride facilities were statistically significant at the 0.01 level in the final model, which produced an R-squared of 0.53 when predicting daily station boardings. The model was used to forecast ridership on a new rail line in Charlotte-Mecklenburg, North Carolina (Parsons Brinkerhoff, 1996).

Another study by Kuby et al. (2004), used data from nine different light-rail systems in the U.S. to forecast station ridership. Variables such as population, employment, and renters within walking distance to the stations, number of park-and-ride spaces, transit connections, the use of heating degree days, and others had the predicted positive or negative coefficients. The model resulted in an R-squared of 0.72 with use of a dependent variable of average weekday boardings.

In addition to light and heavy rail ridership modeling, DRMs have been implemented for bus transit as well. Chu (2004) developed a DRM with Poisson regression for bus routes in Jacksonville, Florida predicting weekday boardings. Socioeconomic variable such as household income, jobs, households with no vehicles, age, sex, race and ethnicity were statistically significant in the model. Chu (2004) also included a transit level of service within a one minute walk and the number of transit stops within a two to five minute walk, as well as a pedestrian factor variable that were all statistically significant in the final model.

Cervero (2006) presented DRMs as an off-line sketch-planning tool that has several advantages over using traditionally four-step travel models in the context of smart growth. This work is particularly significant in the literature because it emphasizes the applicability of DRMs to TOD. Cervero (2006) makes a case for using DRMs to model and evaluate several alternatives by using three examples: greater Charlotte, San Francisco Bay Area exurbs, and south St. Louis County. Models for each of the three examples were statistically significant, and were used to describe the ridership to population density elasticity. Ridership to density elasticities in these models show the impact of increasing (or decreasing) population densities on transit ridership. The model had elasticities of 0.192 on a national scale, 0.233 in the San Fracisco Bay Area, and 0.145 in metropolitan St. Louis. The elasticities are interpreted as, in the case of the San Francisco Bay Area, a 23.3% increase in ridership for every 10% increase in population density. By applying the model to TOD scenarios that doubled the densities around some of the stations in the San Francisco Bay Area, Cervero (2006) estimated that daily ridership would increase between 11 and 17% compared to trend-line forecasts.

Cervero et al. (2010) built a DRM for bus rapid transit (BRT) in Los Angeles County with results encouraging the use of DRMs for BRT ridership in the future. The researchers used a total of 69 BRT stops in the model, with the purpose of predicting the ridership of an improvement of services from mixed traffic to exclusive lanes for the Rapid Blue 3 line operated by the Santa Monica Big Blue Bus. Ordinary Least Squares

(OLS) regression, as well as a Hierarchical Linear Modeling (HLM) was applied to account for the nested nature of the bus stops within bus lines. The OLS model yielded a better statistical fit, with a very high R-squared of 0.952 and all coefficients with the expected signs (Cervero et al., 2010). The results showed that service frequencies and high connectivity to other modes of travel had two of the strongest relationships with daily boardings in the best model. Population density within a 0.5 mile buffer from the stations was the most influential neighborhood characteristic. With all other variables held constant, doubling the population around a single BRT stop within the 0.5 mile distance from 5,000 to 10,000 is estimated to increase daily BRT boardings by 170 (Cervero et al., 2010). Adding an interactive variable modified this relationship, estimating that if the same bus stop had dedicated lane service that employment densities would further increase boardings.

Gutierrez et al. (2011) showed that the average number of boardings in the Madrid Metro network could be refined with a greater level of detail for transit catchment areas. By using distance-decay and network distance to define transit catchment areas, Gutierrez et al. (2011) was able to show pedestrian accessibility to a greater detail than in previous applications of DRMs. Variables that turned out to be highly significant are centrality of the station within the network, employment in commercial and educational sectors, worker and foreigner population groups, the number of transit connections, and land-use mix. The authors note that the variable for no car households was most likely not significant because the locations of the stations are already so dense that having a car was not an advantage. They also speculate that the street density variable was not significant because it is most likely built in to the network calculated service areas.

One issue that Gutierrez et al. (2011) encountered, however, was the poor results from stations that had relatively high intermodal access. In other words, the transit stations in Madrid that experienced ridership from modes other than pedestrian diluted the relationship of the variables in the model to station ridership. The researchers worked around this problem by removing the stations that had intermodal access, using only the stations that had no feeder modes. The authors write that stations with high ridership from a variety of access modes may be better analyzed with a regional approach such as a four-step travel demand model. The elimination of the intermodal stations significantly increased the model's predictive power, indicating that these stations may belong in their own model.

Liu et al. (2014) demonstrates the value of DRMs in the context of smart growth. Focusing on the potential of using DRM results to guide land-use policies, this research includes data from 117 rail stations from light rail and commuter rail in the Baltimore and Washington D.C. metro area. The results indicate that light rail behaves differently than commuter rail. Light rail resulted in statistical significance of employment within 0.5 miles, service level, feeder bus connectivity, distance to CBD, and terminal stations variables. For commuter rail on the other hand, only feeder bus connectivity was significant. For increasing light rail ridership the model shows that employment is the most significant predictor, resulting in 1.6% increase in boardings with every 1% increase in jobs. The authors conclude that DRMs can be useful for rough estimates of station boardings without relying on more complicated regional transportation demand models (Liu et al., 2014).

2.3 Transit catchments

As shown in the previous section, the literature on DRMs has emphasized the beneficial ability to incorporate fine grained resolution of station attributes into the ridership estimation equations. However, the literature also reveals two conflicting schools of thought on how transit catchments are utilized in DRMs. With respect to DRMs, transit catchments are typically the immediately surrounding space of a transit stop or station that is used to define the boundary of relevant characteristics that are incorporated into the model. In praxis, these are 0.5 mile buffers from the station that include characteristics of everything within the boundary and nothing beyond the boundary. They are used as thresholds beyond which socioeconomic characteristics are not included in the model, and are therefore thought of as the boundaries of reasonable access to the station. At the center of the transit area catchment discussion, is the definition and measurement of accessibility. Therefore the definition of the transit catchment area will depend on the accessibility measures used.

Accessibility to a transit feature is widely used as one of the indicators of service performance (Murray, 2001). Transit agencies often strive to maximize accessibility to position themselves more favorably for robust and dependable ridership. One of the most important modes of transportation when connecting to transit is by foot (Biba et al., 2010). In many areas studies have found that most transit riders walk to a transit feature, and nearly all of transit riders reach their destination, by walking. (Biba et al., 2010). In Orange County, California for example, 80% of bus riders walked to the bus from their origin, while 90% of riders walked to their destination (Hsiao et al., 1997). In a large scale survey of over 48,000 transit riders (bus and rail) in the Atlanta metro area, 78.0%

of those riding transit began their trip by walking to the initial transit mode. A complete distribution of mode of access to transit in Atlanta is shown in Figure 2, and further analysis is shown in Chapters 3 and 4. Most transit riders however, are willing to walk only a limited distance to the transit feature. One study showed that the transit rider's willingness to walk to transit drops precipitously after 300 ft. and is almost nonexistent past 0.36 miles (Zhao et al. 2003). Another study found that for every 1,640 ft. in extra walking distance the likelihood of transit riders walking to the transit feature drops by 50% (Loutzenheiser, 1997).

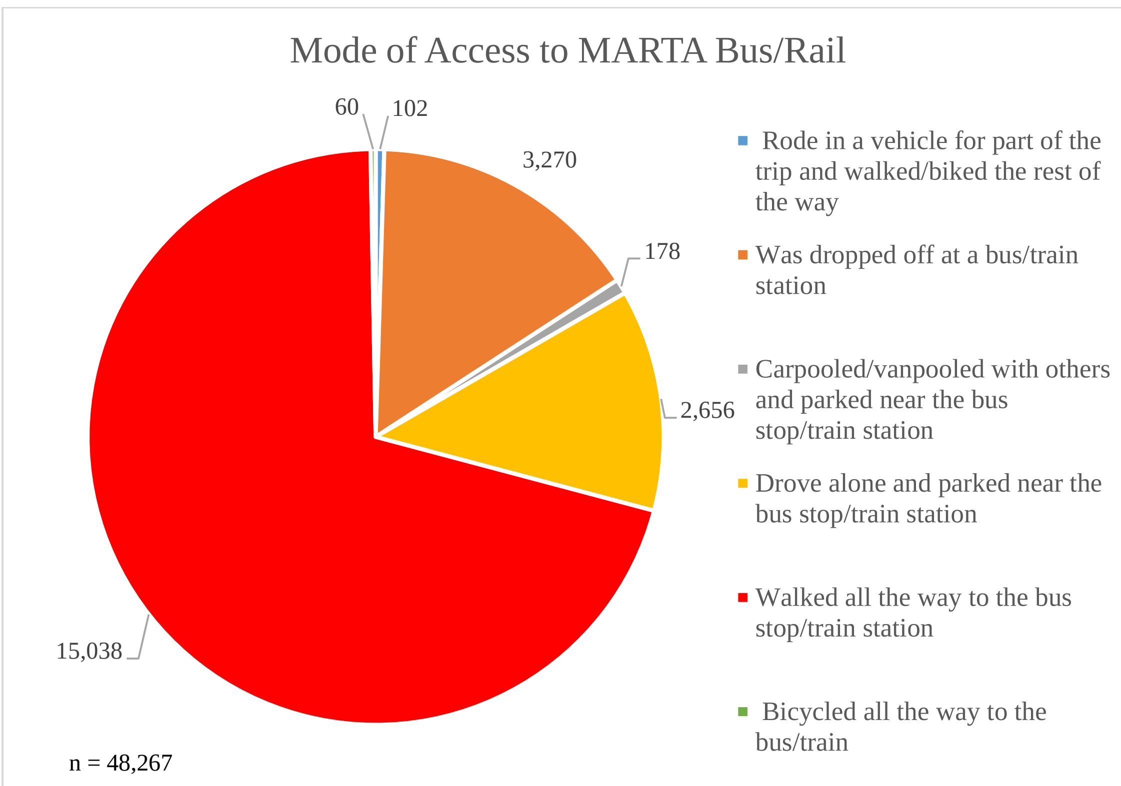

Figure 2: Mode of entry to the MARTA system (bus and rail) from a large-scale ARC survey of riders in Atlanta, GA from October 2009 through January 2010.

The complex and perhaps location specific of transit catchment size has yet to reach a consensus. Some sources are resolute on a 0.25 mile threshold for walking distance to determine accessibility to transit (O'Neil et al., 1992; Hsiao et al., 1997; Philips & Edwards 2002; Biba et al., 2010). The Transit Capacity and Quality of Service Manual (TCQSM) recommends a transit catchment size of 0.5 mile (or about 10 minutes of walking) for rail transit (TCQSM, 2003). Moreover, the shape of transit catchments is often debated, with powerful GIS technology making it easier to construct creative transit catchments that give a greater detail of realistic walkability. Thus, there has been a call in the literature to investigate improvements of measuring accessibility (Biba et al., 2010).

2.3.1 Simple buffer method

The simple buffer method first takes the transit features such as stops, stations, or sometimes routes, and creates a predetermined distance buffer around them (Biba et al., 2010). As previously mentioned, the typical distance is 0.25 or 0.5 miles, although other distances are easily implemented in place of this standard. The buffer method uses what is commonly referred to as straight-line, "as the crow flies", or Euclidian distance, taking advantage of simple straight-line geometry in all directions from the station to create a transit catchment in the shape of a circle. In addition, a multiple ring buffer can be created using a specialized tool in GIS applications, in which analysis can include several buffer-rings to examine population in multi-level approach.

To generate demographic estimates within the transit catchments, census data are downloaded, typically in census tracts, block groups, or blocks, depending on data availability and scale of desired analysis. Population densities are then calculated using the census data and then associated with the 0.25 or 0.5 mile buffer, assuming an even distribution throughout the entire census polygon. An example of the simple buffer method is provided in the top left portion of Figure 3 below, along with the network method and distance decay methods.

Figure 3: Comparison of service areas produced from common methodologies for determining walk access.

The simple buffer method also assumes that population calculated to be within the buffer area will have access to the transit feature (Biba et al., 2010). This is a major flaw in using the buffer method to calculate walking distance, especially in areas that have circuitous road networks and significant pedestrian barriers such as highways, superblocks, or water barriers. In areas with barriers the actual walking distance to get to a transit features could be much greater than the geometric distance. The result in using the buffer method is a consistent overestimation of population within walking distance to transit (Horner & Murray, 2004).

In many of the published papers, the simple buffer method is used for defining the transit catchment area. It has become the standard for planning TODs, loosely based on the reasonable walking distance to a transit station most people are willing to accept (Guerra et al., 2012). Guerra et al. (2012), took a closer look into this assumption specifically for the appropriateness of DRM applications. The work sought to define a clear threshold that may be used as a distance for data inclusion for TOD decision making. With the use of 1,449 transit stations from 21 cities, they used buffers of 0.25 mile, 0.5 mile, 0.75 mile, and 1 mile distances to measure population and jobs. With the inclusion of the variables that are typical to DRMs, Guerra et al. (2012) used OLS regression to show that there was marginal difference in using one threshold over the other. Their work did however, suggest that the use of 0.25 mile buffers for jobs and 0.5 mile buffers for population may be most appropriate for predicting ridership. The authors admonish that the results are "far from definitive", and that further research needs to be completed to ensure the use of correct transit catchment size.

2.3.2 Network ratio method

Another method for estimating the pedestrian accessibility is the network ratio method, which uses a ratio of road networks to more accurately define the percentage of population within walking distance to transit (O'Neill et al., 1992). More specifically, this method calculates the network length of all the roads within the census polygon and all the roads that are in walking distance to the transit feature. By taking the ratio of road length within walking distance to transit to total road network length within the census polygon and applying it to the total population in the census polygon, the assumption of uniform distribution is no longer made. This increases the viability of population estimation because the population is more realistically assumed to be associated with the road network, rather than uniform over space regardless of transportation network (O'Neill et al., 1992).

One serious weakness with this approach, however, is that census boundaries are often directly on streets. It is therefore ambiguous as to which census polygon will be associated with each street, complicating the attribution of population data to the road network and reproducibility of the analysis. Furthermore, like the simple buffer method, the network-ratio method has been shown to also overestimate the population within walking distance to the transit feature (Zhao et al., 2003). This method is still recognized as an improvement over the simple buffer method because of incorporation of the street network and creative usage of street ratios.

2.3.3 Network analysis

GIS technology provides convenient and relatively robust analytical techniques that can be applied to a variety of spatial problems including accessibility. The ArcGIS Network Analyst extension proves to be one of the opportunities for improvement in estimates of population within walking distance to transit. Gutierrez and Garcia-Palomares (2008), calculated coefficients of determination for ridership at transit stations and population estimates using the simple buffer method and network distance for comparison. The results showed that the population estimates calculated by the network distance were better predictors of ridership than the simple straight line buffer method. More work by Gutierrez et al. (2011) demonstrated another successful application of network distance in determination of transit catchment coverage in Madrid.

However, research by Guerra et al. (2012) does not use network distances for transit catchment areas, instead relying on straight line buffer distances. The authors state they use straight line distance because 1) the majority of current direct ridership models use them; and 2) that they add little to no benefit even while increasing the data collection efforts significantly (Guerra et al., 2012). Additionally, when using vehicular roads in creating walking networks, pedestrian paths are not included resulting in smaller catchments in some cases. Conversely, street files do not consider pedestrian barriers, such as streets that lack sidewalks. Simply using street files to construct a pedestrian catchment would also include limited access highways and interstates that are inaccessible to pedestrians. Guerra et al. (2012) suggest that researchers and planners use the most readily available catchment size because the other methods are laborious and counterintuitive to the simplified character of DRMs.

2.3.4 Distance-decay functions

It is well documented that transit patronage decreases as distance to transit increases (Gutierrez et al., 2011). For example, one study in the Netherlands found that those living within a 500-100 m radius of rail stations used the rail 20% less than those living within 500 m of the station (Kreijer & Rietveld, 2000). Research by Untermann (1984) found a decreasing percentage of people willing to walk to transit, with most people willing to walk less than 500 meters but only 10% willing to walk 0.5 miles.

Another study found that while most people were willing to walk to stations that were less than 500 m away, fewer people would walk more decreasing to only 10% at the 0.5 mile distance. Dill et al. (2003) added to this body of knowledge by showing that for a 10% increase in distance there was a corresponding 10% decrease in transit usage.

To account for the decreasing probability of using transit with increasing distances to the transit stop or station, distance decay functions have been measured and applied in a few studies in hopes of creating more accurate transit catchment areas. Gutierrez et al (2011) created a distance decay function from survey data to apply to demographic data in Madrid for a more detailed description of transit accessibility in a DRM. This method involved geocoding origin data linked with corresponding transit station location to model the likelihood of walking to the station from increasing network distances. For each station, 100 meters wide network bands up to 1,500 meters were created. Population and employment data were gathered by transport zones and interpolated within each band using an aerial interpolation method from O'Neil et al. (1992) and Chakraborty and Armstrong (1997). The ratio of riders from each of the network bands to the total population within each band was calculated and incorporated into the distance decay function. The distance decay function was then applied to each of the station's demographic variables creating a distance-decay weighting of the characteristics affecting transit ridership. The results were applied to the regression equation of variables affecting ridership to obtain a distance-decay weighted regression analysis of transit ridership in Madrid, Spain, which improved the predictive power of the model in this study.

Distance decay can readily be applied to either the buffer method or network method with the construction of multiple bands of increasing distance. Figure 2.4.1.1 shows the geographical differences in each of these methods, which yield differences in demographic values used as independent variables in DRMs. While some research points to benefits of using more sophisticated methods such as network distances and distance decay functions (Gutierrez et al., 2011; Biba et al., 2010; Zhao et al., 2003), others argue that simpler methods work similarly well and little is gained with more complicated methods (Cervero et al., 2010; Guerra et al., 2012). Guerra and Cervero (2013), sought to clarify the relationship of transit catchment construction methodology and the ability of the variables to predict ridership in DRMs. In two test cities, Atlanta and San Francisco, varying the size of the transit catchment by increasing or decreasing the radius had little effect on the ability to predict ridership (Guerra & Cervero, 2013).

Due to the lack of consensus in the literature, the most accurate and effective method for determining the size and shape of transit catchments is still unsettled, and warranted for future research (Guerra and Cervero, 2013). However, in light of the research showing marginal differences in transit catchment size and shape, the analysis performed in the following chapters uses the simple buffer method.

CHAPTER 3

OBJECTIVES, DATA AND METHODS

The following study seeks to elucidate the relationship of transit oriented development (TOD) and transit ridership, specifically at Metropolitan Atlanta Rapid Transit Authority (MARTA) rail stations in the Atlanta metropolitan area of Georgia. To model this relationship, the analysis presented in this thesis applies direct ridership model (DRM) methodologies applied in previous research. DRMs have been utilized to reveal valuable insights into the relationship of ridership with environmental characteristics, demographics, and station characteristics. Understanding these relationships can be particularly useful in the context of planning TOD. Currently, MARTA is pursuing several TODs for stations across the metropolitan area in hopes of increasing revenue, ridership, and transportation options for more people. The research presented in this paper builds on the current body of literature with two primary objectives:

1) Investigating the characteristics of transit riders responding to a large-scale travel survey and comparing the results to typical demographic inputs of DRMs
2) Creating a DRM for MARTA's rail station to test relationships of station environment and demographics that are associated with pedestrian-based ridership in the MARTA rail system

The following sections are organized into the data and methods utilized in the effort to assess the three objectives briefly outlined above. The first section (4.1) details the data and methods used for Objective 1, comparing the Atlanta Regional Commission (ARC) travel survey to the demographic data collected from online census sources. The second section (4.2) addresses the chosen data and methods used in creating the DRM for Objective 2.

3.1 Objective 1: Comparison of travel survey to DRM input variables

To help guide the incorporation of variables derived from empirical evidence this research utilizes the Atlanta Regional Commission's 2010 onboard transit survey (ETC Institute, 2010), which is introduced in this chapter and analyzed in Chapter 4. This large-scale survey of travel behavior across the metropolitan region was distributed from October 2009 through January 2010, and included the 20-county planning region shown in Table 1. Transit operators such as MARTA, Cobb Community Transit (CCT), Georgia Regional Transportation Authority (GRTA), Gwinnett County Transit (GCT), Cherokee Area Transportation System (CATS), Hall Area Transit (HAT), and Clayton Transit (CTRAN) were surveyed. The purposes of the survey were to recalibrate the mode choice model and to better understand how transit is being used in the region. Transit patrons were surveyed in an interview style format with either tablet or paper surveys.

Table 1: Counties included in the geographic scope for the survey project.

Clayton	Fayette	Barrow	Hall
Cherokee	Fulton	Bartow	Newton
Cobb	Gwinnett	Carroll	Paulding
DeKalb	Henry	Coweta	Spalding
Douglas	Rockdale	Forsyth	Walton

The largest survey of its kind in the United States, a total of 48,857 on-board surveys were distributed and completed with enough information to be analyzed. A goal of the survey was to obtain a sample of at least 10% from each of the transit systems. The average daily boardings in the MARTA system during the time of the survey was reported to be 403,145. The survey successfully reached a total of 44,006 usable surveys, totaling 10.9% of the average daily boardings (ETC Institute, 2010). For the purposes of analyzing the surveys for this thesis, only surveys of trips accessing MARTA rail stations were utilized, which totaled 21,304 trips. Transit station alighting data were not utilized in this research because transit egress has a different set of restraints that control the mode choice, such has not having vehicles as readily available. The distribution of surveys at each MARTA station is located in Appendix A, including access to rail stations.

The number of riders who arrived at the station via bus were counted as a bus access to the rail mode. The mode to arrive at the bus was disregarded, as the actual arrival mode at the rail station is the focus of the research in this thesis. In the surveys of riders that did not arrive by bus, the following question was used to calculate the distribution of mode access: “How did you get from the place where you started your trip to the very FIRST bus or train you used for this trip?” The respondents had a total of six answers to choose from in response to this question. These include:

- Rode in a vehicle for part of the trip and walked/biked the rest of the way
- Was dropped off at a bus/train station
- Carpooled/vanpooled with others and parked near the bus stop/train station

- Drove alone and parked near the bus stop/train station
- Walked all the way to the bus/train
- Bicycled all the way to the bus/train

In addition to mode of access to the rail station, a deeper investigation into the ARC survey results revealed the characteristics of those riding the transit system. Questions regarding the number of vehicles available, age, income, gender, and race/ethnicity were analyzed to show ridership characteristics and mode of access to the rail station. Differences in characteristics between modes of access are analyzed to gain insight into populations who are more likely to walk to the rail stations.

The results from the ARC survey were then compared to the demographic makeup of transit catchments around each of the rail stations. Comparing the survey data to the census data was done in an effort to gauge the appropriateness of using census data as a proxy for transit riders. American Community Survey (ACS) 2010 five-year estimate data for total population, households, age, sex, and race/ethnicity were downloaded from the American Fact Finder website at the block level. ACS data were specifically chosen because they give granular estimates of demographics and are commonly used in DRMs. In addition, 2010 census data regarding income groups was downloaded from the internet application Social Explorer at the block group level because these data are more suppressed for privacy purposes. Details of the data are shown in Table 2.

Table 2: Demographic variables in transit catchments that were compared to the ARC survey.

Variable	**Source**	**Geography**	**Categories**
Age	American Community Survey (2010 five-year estimates)	Census Block	less than 5, 5 to 9, 10 to 14, 15 to 17, 18 to 19, 20 only, 21only, 22 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 61, 62 to 64, 65 to 66, 67 to 69, 70 to 74, 75 to 79, 80 to 84, 85 years or more
Gender	American Community Survey (2010 five-year estimates)	Census Block	Male, Female
Race	American Community Survey (2010 five-year estimates)	Census Block	White alone, Black or African American, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and other Pacific Islander alone, some other race alone, and two or more races
Ethnicity	American Community Survey (2010 five-year estimates)	Census Block	Not Hispanic or Latino, Hispanic or Latino
Income	Social Explorer (2010 five-year estimates)	Census Block Group	Less than $10,000, $10,000 to $14,999, $15,000 to $19,999, $20,000 to $24,999, $25,000 to $29,999, $30,000 to $34,999, $35,000 to $39,999, $40,000 to $44,999, $45,000 to $49,999, $50,000 to $59,999, $60,000 to $74,999, $75,000 to $99,999, $100,000 to $124,999, $125,000 to $149,999, $150,000 to $199,999, $200,000 or More

For comparing the ARC survey data to the ACS data, a basic transit catchment was created around each of the 38 rail stations using a 0.5 mile straight-line buffer using ESRI's (Environmental Systems Research Institute) ArcMap 10.3 spatial analysis program. Estimates of population demographics around each station were derived from calculating densities of ACS demographics and multiplying the result by the clipped areas of census blocks within the transit catchments.

The straight-line buffer distance of 0.5 miles is displayed cartographically in Figure 4 below. However, to eliminate overlap of transit catchments that would "double-count" populations that are close to multiple stations, Thiessen (Voronoi) polygons were drawn around the 38 stations and used as boundaries between stations, shown in Figure 5. Creating Thiessen polygons using the 38 stations as points divides the Atlanta area into zones, where every location inside each zone is closest to its respective station. In cases with overlapping transit catchments, the Thiessen polygon defines the boundary where one station's catchment stops and another begins. The Thiessen polygon method assumes that populations that are within a 0.5 mile walking distance to multiple stations will chose the station that has the shortest straight line distance. The following section describing Objective 2 utilizes this methodology to clip demographic and environmental variables to transit catchments, obtaining values for each of the variables to incorporate into the DRM.

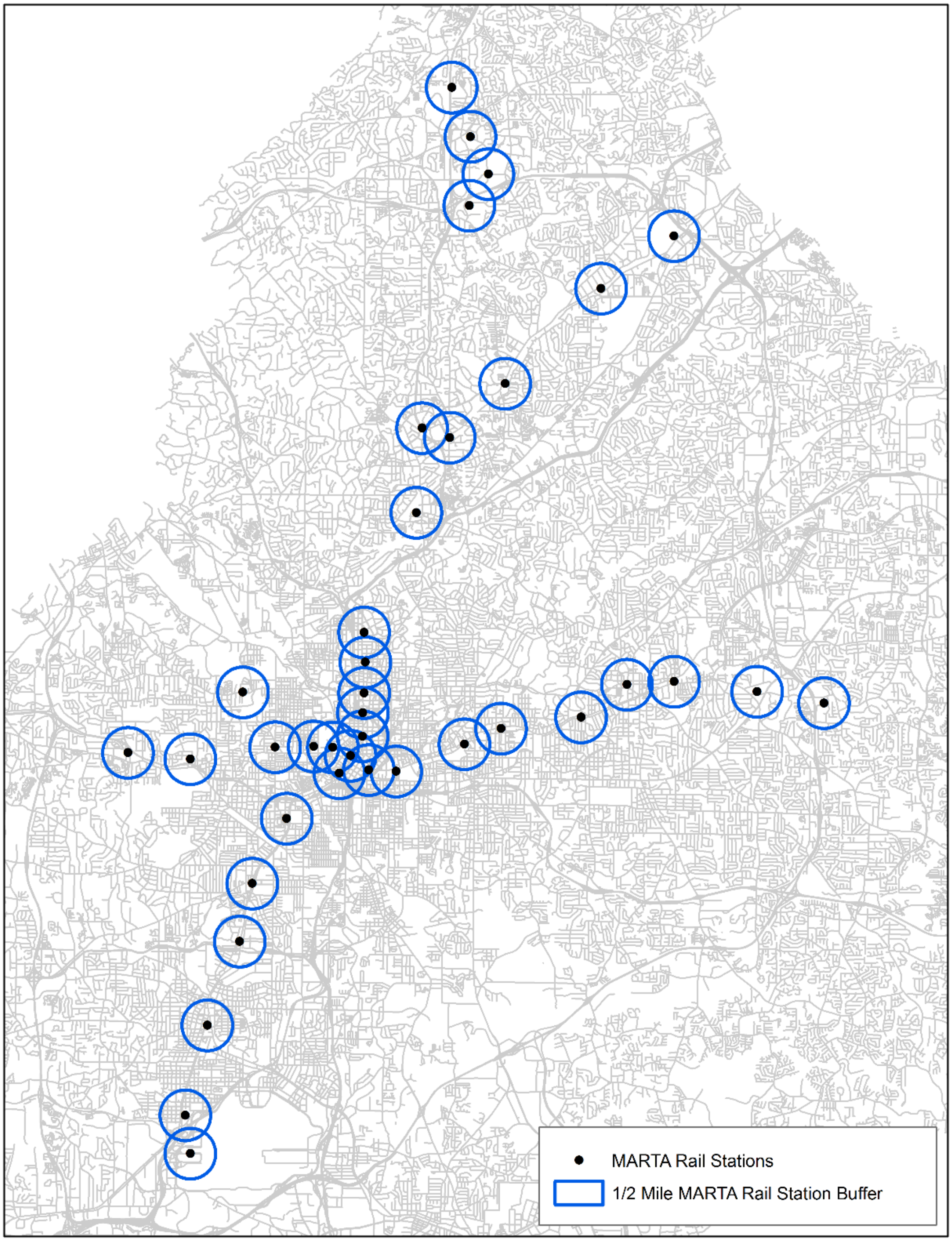

Figure 4: Map of MARTA rail stations with associated 0.5 mile buffer.

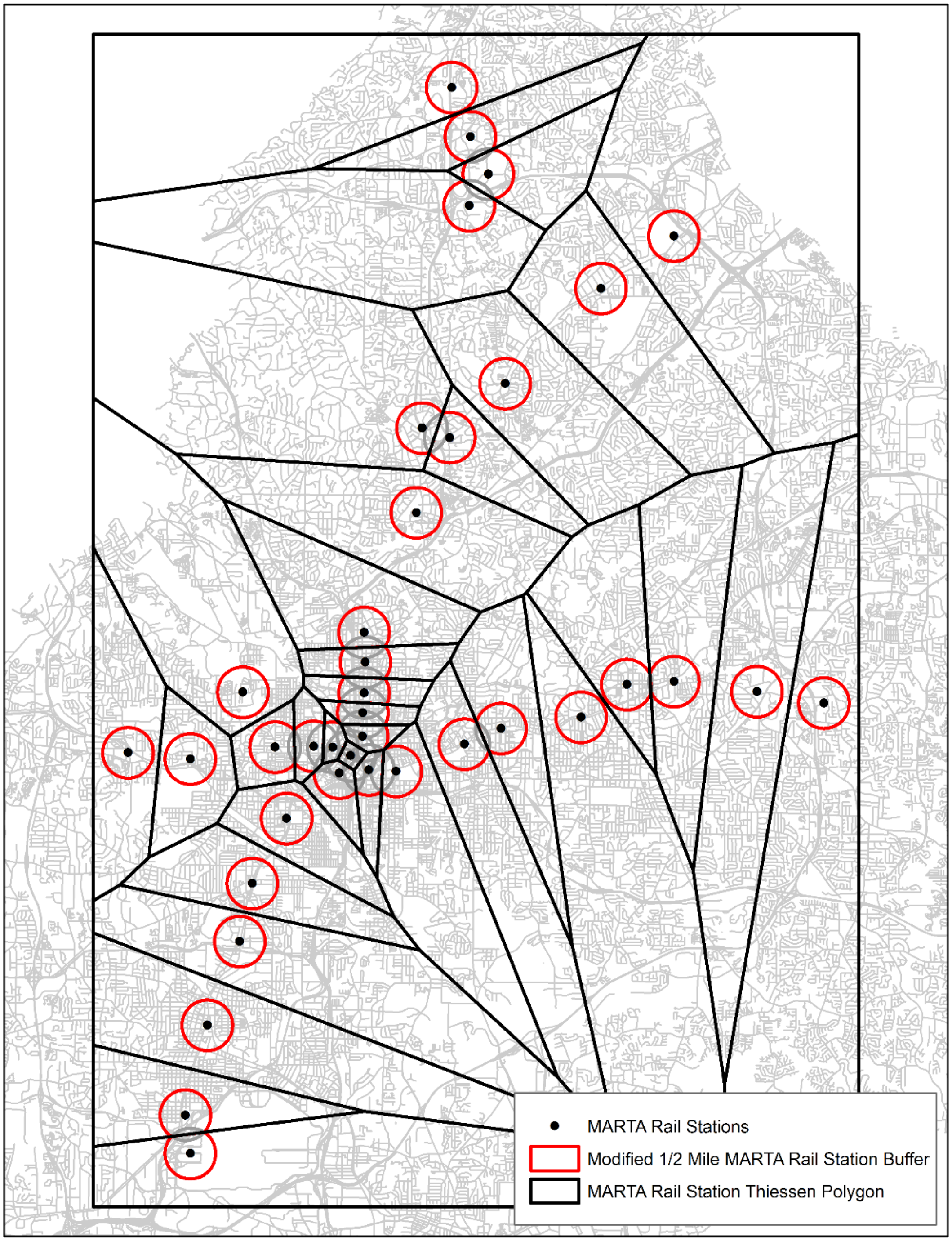

Figure 5: Map of MARTA rail stations with associated Thiessen polygons used to modify the 0.5 mile transit catchment buffers.

3.2 Objective 2: DRM creation for MARTA rail system

For the second objective, this paper seeks to construct a DRM for the MARTA rail system using ordinary least squares (OLS) regression with variables from a variety of sources to predict transit ridership. OLS was chosen based on the existing literature showing its effectiveness in DRMs based on the accuracy, speed, flexibility, and overall ease of interpretation. However, this paper departs from the existing literature in an attempt to focus only on the pedestrian ridership of the rail stations.

3.2.1 DRM dependent variable

The majority of DRM research uses total ridership at each station as the dependent variable in the model. This often involves using small-scale station characteristics and demographics of transit catchments that are typically drawn with the simple 0.5 mile buffer method around the station and using them to estimate total boardings at the station, regardless of access mode to the station. Demographics of riders who access the station from areas outside of the transit catchment are therefore not included in the model. This method could present misleading results in DRMs that do not have sufficient variables that account for ridership from buses and park and rider users, who are more likely to have accessed the station from greater distances than those who have walked. Gutierrez et al. (2011) showed that removal of intermodal stations in Madrid improved the DRM results significantly. Because of the small number of rail stations in the MARTA system (38 total), the removal of these stations was not preferred. Instead, this paper utilizes survey results to estimate the number of boardings at each rail station accessed on foot. Therefore the variables of the station environment will be measured with the pedestrian ridership rather than total ridership at the station.

Moreover, this method intends to clarify the relationship of station environment and elements of TOD that affect pedestrian access without access from bus or personal vehicle confounding this specific relationship.

In addition to guiding the choice of variables in the model, the ARC travel survey was used to estimate pedestrian access to rail stations, the dependent variable in the model. To estimate the number of pedestrians accessing each of the stations, the total number of those indicating walk access to the station was compared to the total number of those accessing the station. The percentage of pedestrian access riders was then calculated and multiplied by the average weekday boardings at each of the MARTA rail stations from October 2009 through January 2010. This resulted in estimates of pedestrian ridership per stations (weekday riders entering the MARTA rail station by foot), which was used as the dependent variable in the DRM.

3.2.2 DRM independent variables

The DRM includes a variety of independent variables that have been shown to affect transit ridership in other studies. These include demographic variables, station environmental variables, and service variables. To define the area from which the demographic and environmental variables are derived, a transit catchment size of 0.5 miles was computed following the methodology outlined at the end of section 3.1. The demographic variables included in the model come from the online census sources revealed to show the greatest impact from the ARC on-board survey analysis. In addition, the model incorporates the Claritas Parcel Level Employment Database from 2010 to obtain jobs by industry. This was obtained at the census block level with 6-digit North American Industry Classification System (NAICS) codes. The number of possible codes totaled 19,280, which was summarized into 21 two-digit codes for simplicity: Agriculture, Forestry, Fishing and Hunting; Mining, Quarrying, and Oil and Gas Extraction; Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; Administrative and Support and Waste Management and Remediation Services; Educational Services; Health Care and Social Assistance; Arts, Entertainment, and Recreation; Accommodation and Food Services; Other Services (except Public Administration); Public Administration; and Non-classifiable Establishments.

Station environment variables tested in the DRM come from a variety of sources. Land-use mix was calculated following the Liu et al. (2014) methodology, in which the percentage of each use was measured in creation of a land-use mix index, written here as LUMI. The LUMI uses the square footage of commercial, residential, and industrial floor area to show how evenly distributed the different land-uses are in values ranging from 0 (where the area is covered by only one type of land use) to 1 (where land-use is in perfect even distribution). The LUMI was calculated using the equation:

$$\mathrm{LUMI} = ((-1)/\ln n) * \sum_{i=1}^{n} p_i \ln p_i$$

where n is the total number of land-use types, and p_i is the percentage of land-use type i of the total land area. Values closer to 1 (heterogeneous land distribution), are expected to encourage transit patronage while values closer to 0 (homogenous land distribution) are expected to discourage transit patronage.

Several other station area environment variables were used as candidate variables as well, including street density, intersection density, end-of-line stations, and number of parking spaces at MARTA stations. Calculating the street density of an area is sometimes used as a proxy for connectivity (Dill, 2004). Greater values indicate a greater number of streets and therefore a greater likelihood of pedestrian connectivity. Here, the street density was determined by calculating the length of surface street network in each of the transit catchments in ArcMap 10.3. Topologically Integrated Geographic Encoding and Referencing (TIGER) 2010 street files were downloaded and the limited access freeways were deleted from the dataset to avoid including pedestrian inaccessible roads from the value. Once freeways were removed, the length of roads was calculated and divided by the area of the transit catchment, yielding a street density variable in miles of streets per square mile. In addition to street density, the number of intersections were calculated by counting nodes in the street network. The number of intersections can be used as a measure of street connectivity, as shorter blocks and a greater number of intersections can lead to shorter overall walking paths and greater accessibility for pedestrians.

The number of parking spaces has also been used in DRMs to help prognosticate ridership at the station level. While parking utilization would be a more appropriate indicator of the number of riders that access the rail stations via personal automobile, it may still indicate the general attractiveness of a station. A station that is tailored for auto access may have a negative relationship with pedestrian ridership, and determining such would also be valuable. Unfortunately, parking utilization was not available during the study period. Instead, the total number of parking spaces was included as a candidate variable in hopes that the number of parking spaces available is generally indicative of an auto-focused rail station, with potentially a negative effect on pedestrian boardings.

Service characteristics variables may help predict the ridership at transit stations as well. Transit agencies will often respond to large demand with increased service to meet the needs of riders. Conversely, frequent service may entice greater numbers of riders to utilize a particular station. However, as noted by Taylor and Fink (2003) and then by Gutierrez et al. (2011), introducing service frequency characteristics into a DRM may produce endogeneity problems between the service supply and service demand. Moreover, the inclusion of a bus feeder variable would be a more appropriate variable for a DRM considering total ridership rather than ridership only from pedestrians. A bus feeder variable may, however, depress pedestrian ridership because riders may ride the bus to the rail station rather than walking, especially in the case of a free transfer to rail as in the MARTA system. For these reasons, bus and rail service frequency characteristics were not included as candidate variables in this particular model. An end of line dummy candidate variable was, however, included because the nature of these stations is often different than other stations in form and function.

3.2.3 Decision tree analysis

Binary recursive partitioning (BRP), also known as decision tree analysis or Classification and Regression Tree (CART) analysis, was performed to gain further insight into the candidate variables influence on pedestrian ridership. BRP is a statistical method for predicting a response variable based on values of predictor variables (Merkle & Shaffer, 2011). The output of BRP is a decision tree that shows the predictor variables in a hierarchy, with the single most important predictor at the top of the tree. The process works by testing all of the independent variables for the strongest association with the dependent variable and responds by placing it at the top of the tree, from which other predictor variables are grown. BRP calculates a threshold for the most important predictor variable, and two stems from the variable are created. One of the resulting stems signifies values that meet the conditions of the most important predictor variables (true stem), and the other stem signifies values that do not meet the conditions of the most important predictor variable (false stem). Each stem (also referred to as a split) leads to either another predictor variable (known as a node) or a value referring to the response variable value (known as a terminal node). BRP works to reveal interactions between variables, as decision trees with multiple splits indicate situations where the lower predictor variables are subsets that meet the higher predictor variables categorizations as well.

Merkle and Shaffer, (2011), outline a simple example of housing prices where BRP can be useful. If one used housing prices as a response variable, BRP may identify a "number of bedrooms" predictor variable to be the most influential predictor of price. BRP would place the "number of bedrooms" predictor variable at the top of the tree and split it into two groups stemming down from the top of the tree. For example, perhaps having three or more bedrooms would indicate higher prices, and less than three would indicate lower prices. BRP would then identify the next most influential predictor variable, which may be the distance from the city, represented as a subgroup from one of the stems.

For the purposes of creating a DRM, BRP is used not only to help define the most influential predictor variables to test in the OLS model, but it also shows the interaction of predictor variables for use in the model. Once the most influential variables are produced, the splits in the BRP will show the most accurate predictor variable in each of the subsets of the predictor variables. The tree can then be used to help identify which candidate variables should be used, and how the variables interact with each other to produce greater ridership through variable combinations and break points.

3.2.4 Ordinary least squares regression

Many DRMs are constructed using OLS regression, which was the chosen method for determining the functional form of the model. OLS regression has been shown to model transit ridership with relative success in a number of examples described in Chapter 2. Moreover, the ease of interpretation of OLS provides additional support in use of this type of model in forecasting transit ridership. The variables detailed in the previous section were imported into Statistical Package for the Social Sciences (SPSS) version 22. A correlation matrix was performed to show how related each of the variables are to the dependent variable, pedestrian ridership. Variables that were not well correlated with pedestrian ridership were excluded from further analysis, while variables that performed well in explaining the variation in pedestrian ridership were retained. Further reduction in candidate variables was performed by removing variables that were highly correlated, to lower the likelihood that the final model would suffer from multicollinearity. Additionally, residuals were plotted to investigate the appropriateness of using a linear model as a method to forecast pedestrian boardings.

CHAPTER 4

ANALYTICAL RESULTS

An investigation into the Atlanta Regional Commission (ARC) survey variables that could influence mode of access to the rail stations is presented in section 4.1. Specifically, the mode of access, the number of vehicles available, income, age, gender, ethnicity, and race are tabulated and analyzed to show the characteristics of rail patrons that walk to the station. Additionally, demographics obtained from census data are compared to the on-board survey data to show how typical variables used for direct ridership model (DRM) input differs from actual riders. Variables that quantify characteristics of the station environment and characteristics are analyzed in section 4.2. Following the results from the three categories of candidate variables are the results from the decision tree analysis, shown in section 4.3. Chapter 4 concludes with a section on the DRM construction and results, revealing the relationship of the variables from the final model and pedestrian ridership at Metropolitan Atlanta Rapid Transit Authority (MARTA) stations.

4.1 Survey Results

First, the ARC survey data were analyzed to glean more information on transit patrons who are more likely to walk to the rail station than arrive by other modes. After the initial investigation into mode, a deeper investigation into how the rider demographics change by mode was executed. Statistics on vehicle ownership, household income, age, gender, ethnicity, and race were calculated and compared to transit catchment demographics to see if the riders represented in the ARC ridership survey were also representative of those living within walking distance of rail stations. Cross tabulations for the survey data help define the differences among transit riders by mode of access as well.

4.1.1 Mode of access

A total of 21,304 respondents to the ARC On-Board Transit Survey indicated the mode of entry to the train station (ETC Institute, 2010). Shown in Figure 6, the most common mode of accessing the rail stations was via walk access, shown by the 8,242 (38.7%) respondents choosing the answer: "Walked all the way to the train station". The second most common mode of accessing the rail stations was arriving by bus, which totaled 7,782 respondents (36.53%) of those surveyed. After bus arrival, the answer "Drove alone and parked near the train station" was the most common, which totaled 2,547 respondents (11.96%). The next most common mode was to be "dropped off at a train station, which totaled 2,452 (11.51%) of the total answers. Next, the response of "Carpooled/vanpooled with others and parked near the bus stop/train station" was tallied 168 (0.79%) times. A total of 65 (0.31%) of the 21,304 respondents answered that they "Rode in a vehicle for part of the trip and walked/biked the rest of the way". The least common answer to this question was that they "Bicycled all the way to the train station", which was only 48 (0.23%).

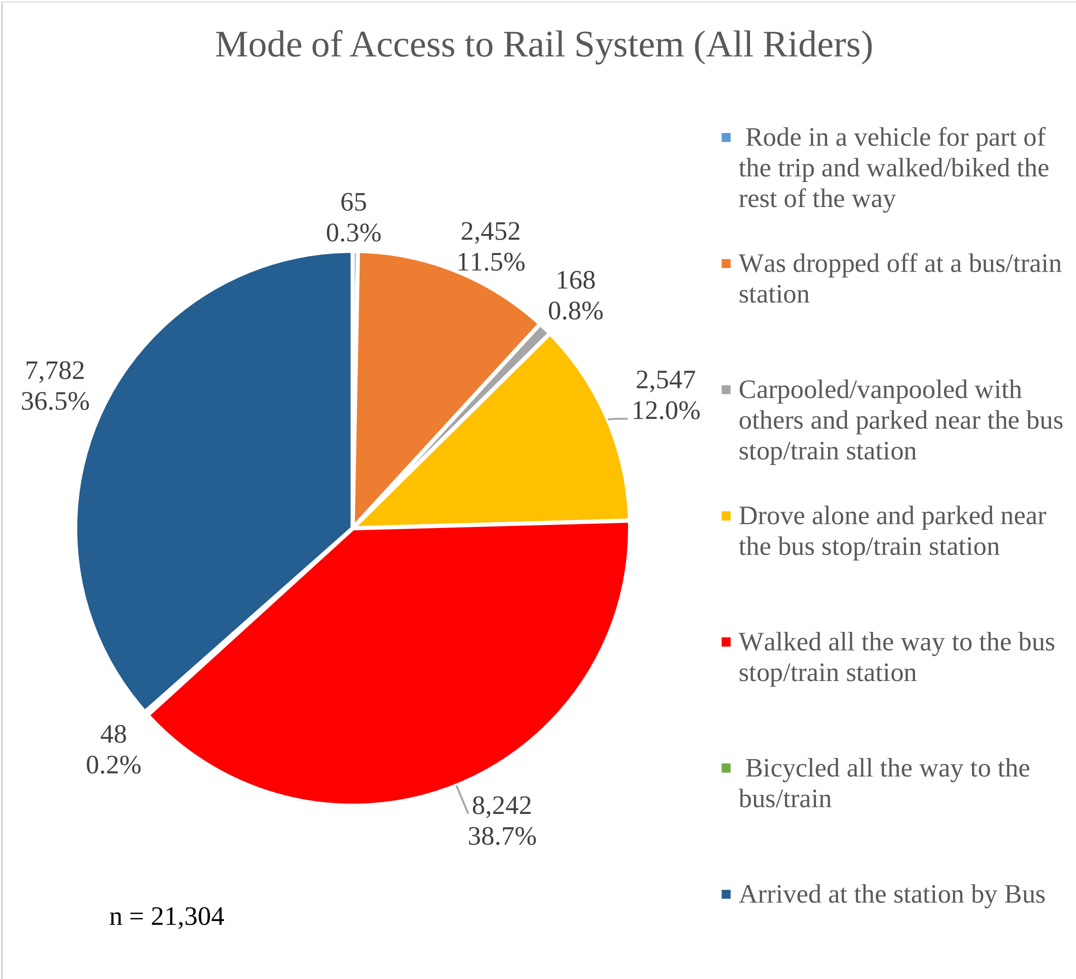

Figure 6: ARC survey results of mode of access for those entering the MARTA rail system.

The distribution of mode at each station varied widely. Some of this variation can be attributed to the tendency of inbound trips to have different modes of access than outbound trips. For instance, a passenger who takes an outbound trip to work from home may be more inclined to drive to the station rather than walk. Conversely, the same passenger returning from work may not have a vehicle to access the rail station on the inbound trip back home, and therefore walks instead. Station surrounded by employment may result in greater walk trips to transit than stations with more residences nearby. With this in mind, the greatest percentage of respondents who reported walking all the way to the train station, in order, were reported at Peachtree Center (92.8%), Dome/GWCC/Philips/CNN (91.7%), GA State (87.9%), Civic Center (86.4%), and North Avenue (78.2%). All of these stations are close to the downtown/midtown areas in Atlanta. Greater pedestrian ridership at these stations is in line with research shown in Chapter 2, which links central business districts (CBDs) and high density to increased ridership at transit stations. Variables that emphasize density are explored and incorporated into the model later in this chapter. Conversely, the lowest percentage of respondents who reported walking all the way to the train station, in order, were reported at Indian Creek (3.2%), Doraville (4.5%), College Park (8.9%), Oakland City (10.3%), and North Springs (11.3%). It should be noted that Indian Creek, Doraville, and North Springs are all end-of-line stations, and that four of the seven stations with the lowest walk percentage are end-of-line stations. Ridership research shows that end-of-line stations typically have a greater share of car ridership relative to other access modes. It is therefore suggested to include an end-of-line dummy variable in the DRM to potentially account for such a difference.

4.1.2 Number of vehicles available

Of those surveyed, 7,087 (33.3%) respondents reported having no available vehicles at their household, 7,299 (34.3%) reported having one vehicle available, 4,963 (23.3%) had two vehicles available, and 1,955 (9.2%) had three or more vehicles available. These values are compared with the vehicle ownership of rail riders who entered the system by foot, shown graphically in Figure 7. Of those that entered the transit system by walking to a rail station, 2,622 (33.8%) indicated that they had 0 vehicles available, 2,844 (34.5%) indicated that they had one vehicle available, 1,980 (24.0%) indicated that they had two vehicles available, and 796 (9.7%) indicated that they had three or more vehicles available. In all, about two thirds (66.7%) of respondents had at least one vehicle available and the remaining one-third had no vehicles available. The percentages of vehicle ownership are surprisingly similar for riders that entered the system by walking and the rail riders as a whole. This similarity suggests that those living within walking distance to a rail station and choose to walk to take rail have the same likelihood of vehicle ownership as rail patrons as a whole.

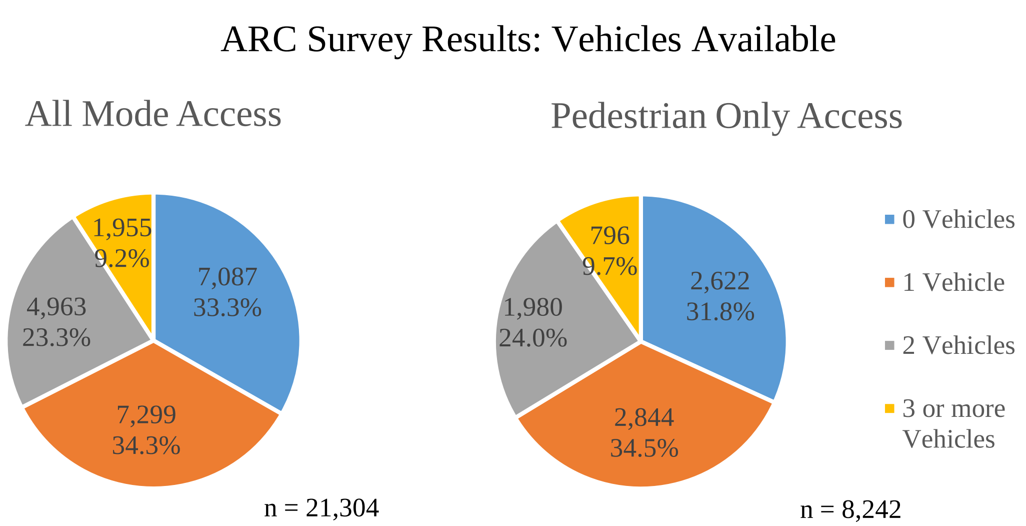

Figure 7: Vehicles available to the households of those entering the MARTA system by all modes (left) and by pedestrian access only (right).

To test this relationship further, contingency tables have been calculated by cross-classifying (or cross-tabulating) the data to see if vehicle ownership differs by mode of rail station access. Table 3 shows the observed counts and the expected counts for the mode of access to the rail station and the number of vehicles available. Table 4 shows the chi-squared statistic for the test of independence was 2,193.6. Compared to a five percent critical value of only 28.9, the null hypothesis that vehicle ownership and mode of MARTA rail access are independent can be rejected.

Table 3: Cross tabulated mode of access and number of vehicles available.

			Vehicles Available				
			Zero	One	Two	Three or more	Total
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	11.0	22.0	24.0	8.0	65.0
		Expected Count	21.6	22.3	15.1	6.0	65.0
	Was dropped off at a bus/train station	Count	662.0	923.0	646.0	221.0	2,452.0
		Expected Count	815.7	840.1	571.2	225.0	2,452.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	39.0	60.0	52.0	17.0	168.0
		Expected Count	55.9	57.6	39.1	15.4	168.0
	Drove alone and parked near the bus stop/train station	Count	54.0	1,014.0	964.0	515.0	2,547.0
		Expected Count	847.3	872.6	593.4	233.7	2,547.0
	Walked all the way to the bus stop/train station	Count	2,622.0	2,844.0	1,980.0	796.0	8,242.0
		Expected Count	2,741.8	2,823.8	1,920.1	756.3	8,242.0
	Bicycled all the way to the bus/train	Count	19.0	14.0	13.0	2.0	48.0
		Expected Count	16.0	16.4	11.2	4.4	48.0
	Rode the bus to the station	Count	3,680.0	2,422.0	1,284.0	396.0	7,782.0
		Expected Count	2,588.8	2,666.2	1,812.9	714.1	7,782.0
Total		Count	7,087.0	7,299.0	4,963.0	1,955.0	21,304.0
		Expected Count	7,087.0	7,299.0	4,963.0	1,955.0	21,304.0

Table 4: Chi-Squared Test for independence for the mode of access to the rail stations and the number of vehicles available.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,193.589[a]	18	0.000
Likelihood Ratio	2,608.883	18	0.000
N of Valid Cases	21,304		

a. 1 cells (3.6%) have expected count less than 5. (the minimum expected count is 4.40)

H0: The mode of MARTA rail access and the number of vehicles available are independent.
Ha: The mode of MARTA rail access and the number of vehicles available are not independent.

Although many of the observed versus the expected counts are relatively close, there are some patterns of large differences. For those who drove alone to the station, the observed counts for zero vehicles are much lower than the expected. Moreover, the observed counts for one, two, and three or more are much higher than the expected counts. This relationship is straightforward, indicating that riders who drive alone to the station are very likely to have one or more vehicles available to them. However, an interesting trend is that the proportional difference of observed and expected grows as the number of vehicles available increases after zero (16.2%, 62.4%, and 120.3% for one, two, and three or more vehicles, respectively). This may indicate that compared to other modes, the likelihood of driving alone to a MARTA station becomes more likely as the number of vehicles available to that individual increases. The relationship of having a greater number of vehicles to driving alone is also very likely related to having a greater income.

Another source of disparity of observed and expected is the "Rode the bus to the station" category, where the number of people with zero cars is much higher than expected, and the number of people with two and three or more cars is much less than expected. This indicates another intuitive relationship, where those who ride the bus are less likely to have access to vehicles. Interestingly, the population that walked to the rail station did not exhibit these large disparities, shown in Figure 6, with remarkably similar vehicle ownership to the entire sample of MARTA rail patrons. For the purposes of constructing a DRM for pedestrian based access, auto-ownership may provide less insight on whether someone is walking to the station or not than other station variables. Auto ownership instead determine if a rider outside of the transit catchment drives or takes a bus to the station.

Figure 8 graphically shows the distribution of mode of access to rail distribution by the number of vehicles available. Of those that had zero vehicles available, the most common mode of entering the rail system was by bus, with 3,680 (51.9%). This represents a large departure from an expected count of only 2,558.8. The next most common was by walking to the rail station, at 37.0%, which was very close to the expected count of 2,741.8. The remaining counts of those with zero cars available were all within reasonable ranges to the expected counts considering the sample size. For rail riders who had one car available, the likelihood of reaching the station via bus decreased dramatically to 33.2%. This was replaced by small increases in walking and being dropped off, with the majority of the difference being made up with those that drove alone, which made up 13.9%. Moreover, this trend continues as more vehicles are available. A smaller percentage of those with more cars are taking the bus to get to the station, and instead are driving alone to get the stations.

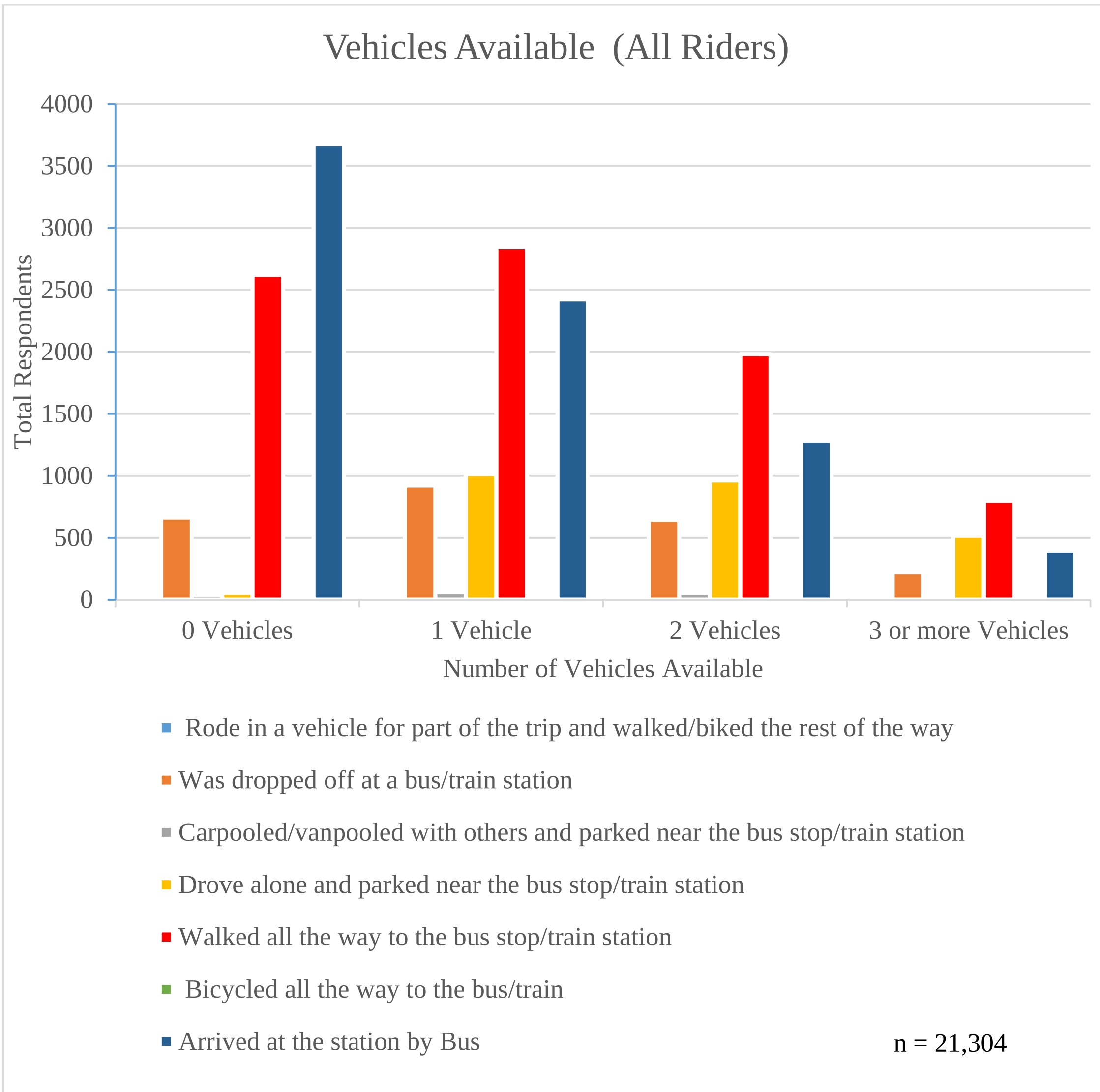

Figure 8: Number of vehicles available by mode of entry to the rail stations.

At the station level, four out of the top five stations with the greatest percentage of respondents with zero cars available were on the Blue and Green lines, west of Five

Points Center Station. The stations with the greatest number of zero cars available were Ashby (54.1%), Oakland City (53.0%), West Lake (52.3%), Vine City (50.3%), and Bankhead (49.00%). Conversely, stations with the greatest percentage of at least one car available are North Springs (80.8%), Edgewood/Candler (80.4%), Peachtree Center (80.0%), Indian Creek (78.1%), and Medical Center (77.5%). This shows a regional trend of higher auto availability along the two northern lines and lower auto availability along the south and west lines, which is likely correlated to income.

4.1.3 Income

A total of 20,500 survey respondents indicated the annual income of their household. Table 2 in Chapter 4 shows the categories available to the respondents, while Figure 9 below shows the distribution of results. The mode for this question is $30,000-$39,999 (15.4% of total), followed by $20,000-$29,999 (14.2% of total), below $5,000 (12.8% of total), and $10,000-$19,999 (10.8% of total). The three highest income categories ($75,000-$99,999; $100,000-$119,999; and $120,000 or more) were also the least common responses to the survey, with only 6.2%, 3.4%, and 5.5%, respectively. Overall, most respondents reported incomes that fall in the middle to low income categories, with a slight positive skew towards the higher income categories. A relatively large percentage of riders responded in the lowest income category.

At the station level, the category of $30,000-$39,999 was the most common for 18 out of the 38 rail stations. The station with the greatest mode was Peachtree Center, with 14.5% indicating a household income of $120,000 or more. North Springs had the second greatest mode, with 13.0% indicating $75,000-$99,999. There were eight stations with the lowest category of below $5,000. These were West Lake (24.0%), Bankhead

(23.0%), Avondale (20.7%), Ashby (20.5%), Oakland City (18.9%), Lakewood/Fort McPherson (16.7%), GA State (14.8%), and Midtown (13.1%).

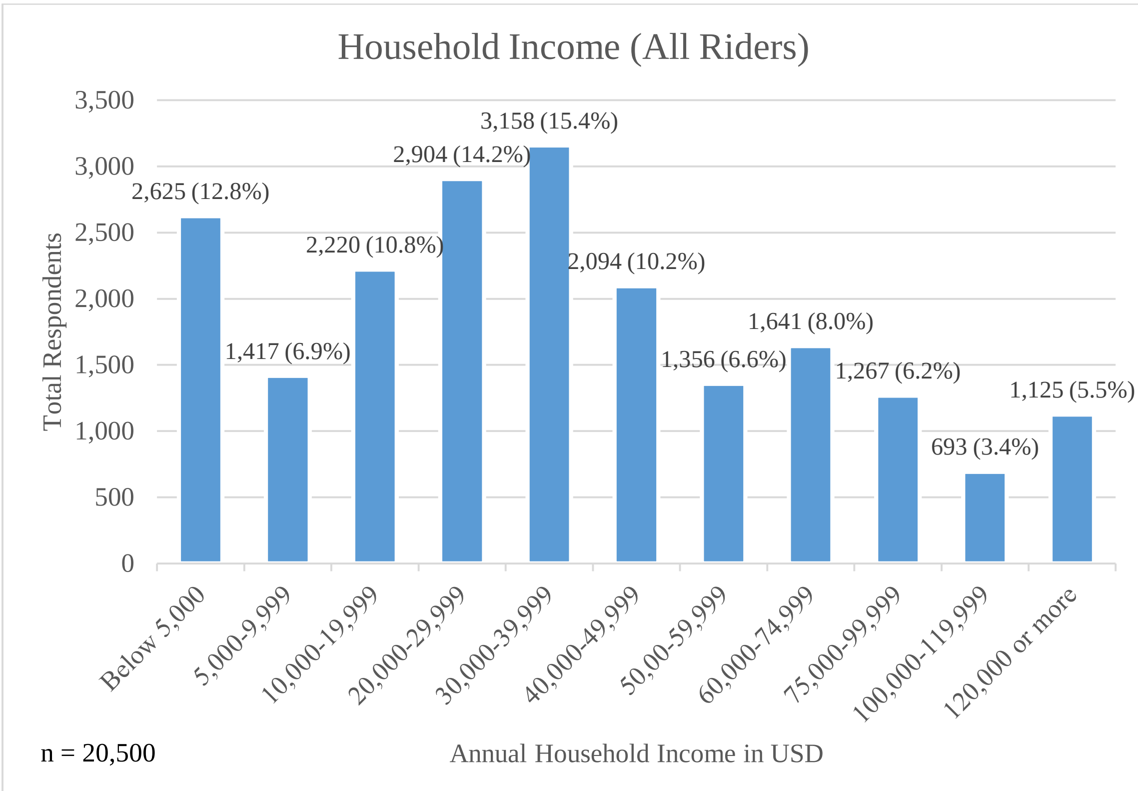

Figure 9: ARC survey results of household income of those entering the MARTA rail system.

Similar to the previous section concerning vehicles available and mode of access, cross tabulations and chi-squared statistics were calculated for household income and mode of access to rail. Because of the large number of categories of responses (11 total), the cross tabulations of observed counts and expected counts are separated into two tables, with incomes less than $50,000 in Table 5 and incomes of $50,000 or more in

Table 6. Table 7 shows the Chi-Square Test for Independence, which reveals a chi-squared statistic of 2,004.2. When compared to the critical value of 78.8, the null hypothesis that income and mode of MARTA rail access are independent can be rejected.

Table 5: Cross tabulation of mode of access and income (below $5,000 to $40,000-$49,999).

			Income					
			Below $5,000	$5,000-$9,999	$10,000-$19,999	$20,000-$29,999	$30,000-$39,999	$40,000-$49,999
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	12.0	1.0	13.0	6.0	3.0	10.0
		Expected Count	8.1	4.4	6.8	8.9	9.7	6.4
	Was dropped off at a bus/train station	Count	306.0	164.0	224.0	319.0	412.0	243.0
		Expected Count	302.7	162.9	255.1	333.7.2	362.9	241.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	17.0	7.0	14.0	14.0	19.0	21.0
		Expected Count	20.4	11.0	17.2	22.5	24.5	16.5
	Drove alone and parked near the bus stop/train station	Count	166.0	62.0	115.0	199.0	314.0	309.0
		Expected Count	317.3	171.3	268.3	351.0	381.7	250.3
	Walked all the way to the bus stop/train station	Count	832.0	442.0	774.0	1,053.0	1,162.0	813.0
		Expected Count	1,007.1	543.6	851.7	1,114.1	1,211.6	810.1
	Bicycled all the way to the bus/train	Count	0.0	4.0	7.0	9.0	4.0	8.0
		Expected Count	6.0	3.2	5.1	6.7	7.2	4.7
	Rode the bus to the station	Count	1,292.0	737.0	1,073.0	1,304.0	1,244.0	690.0
		Expected Count	964.5	520.6	815.7	1,067.0	1,160.3	764.9
Total		Count	2,625.0	1,417.0	2,220.0	2,904.0	3,158.0	2,094.0
		Expected Count	2,625.0	1,417.0	2,220.0	2,904.0	3,158.0	2,094.0

Table 6: Cross tabulated mode of access and income ($50,000-$59,000 to $120,000 or more).

			Income					
			$50,000-$59,999	$60,000-$74,000	$75,000-$99,999	$100,000-$119,000	$120,000 or more	Total
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	1.0	6.0	5.0	1.0	5.0	63.0
		Expected Count	4.2	5.0	3.9	2.1	3.4	63.0
	Was dropped off at a bus/train station	Count	150.0	196.0	157.0	87.0	98.0	2,356.0
		Expected Count	155.8	188.6	145.6	79.8	129.5	2,356.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	15.0	13.0	21.0	4.0	14.0	159.0
		Expected Count	10.5	12.7	9.8	5.5	8.9	159.0
	Drove alone and parked near the bus stop/train station	Count	223.0	336.0	309.0	182.0	263.0	2,478.0
		Expected Count	163.9	198.4	153.2	82.9	134.5	2,478.0
	Walked all the way to the bus stop/train station	Count	559.0	718.0	564.0	345.0	603.0	7,865.0
		Expected Count	520.2	629.6	486.1	268.1	435.2	7,865.0
	Bicycled all the way to the bus/train	Count	4.0	4.0	3.0	1.0	3.0	47.0
		Expected Count	3.1	3.8	2.9	1.6	2.5	47.0
	Rode the bus to the station	Count	404.0	368.0	208.0	73.0	139.0	7,532.0
		Expected Count	498.2	602.9	465.5	253.1	410.9	7,532.0
Total		Count	1,356.0	1,641.0	1,267.0	693.0	1,125.0	25,000.0
		Expected Count	1,356.0	1,641.0	1,267.0	693.0	1,125.0	25,000.0

Table 7: Chi-Squared Test for Independence for the mode of access to rail stations and income.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,004.199[a]	60	0.000
Likelihood Ratio	2,103.66	60	0.000
N of Valid Cases	20,500		

a. 12 cells (15.6%) have expected count less than 5. (the minimum expected count is 1.59)

H0: The mode of MARTA rail access and the number of vehicles available are independent.
Ha: The mode of MARTA rail access and the number of vehicles available are not independent.

In support of the cross-tabulations, Figure 10 shows the mode of station access and the annual household income of respondents graphically. Of respondents that walked to the rail station, the most were from the $30,000-$39,999 (1,162) category, followed by the $20,000-$29,999 (1,053) category, and the below $5,000 category (832). The higher income categories were consistently less populated than the lower income categories. However, when looking at the mode of access share within each of the income categories, the highest income categories had the greatest percentage of people that walked to the rail station compared to other modes of access. For example, 53.6% of those with household incomes of $120,000 or more walked to the stations instead of accessing rail by other modes, compared to only 31.2% of those in the $5,000-$9,999 income category that walked to the rail station. In addition, Table 5 and Table 6 show that the observed count for those walking to rail is consistently lower than the expected count in the lowest five income categories and consistently higher than the expected count in the highest income categories. Transit riders in the lower income categories appear to be less likely to walk to rail than the higher income categories.

The same trend occurs for the driving alone to the station category and income. More than 20% of those in each of the highest four income categories drove alone to the station, while the lowest five income categories each had less than 10% access the rail via driving alone. Again, the higher income categories had an overrepresentation of driving alone to the station, while the lower income categories had an underrepresentation of driving alone to the station. This suggests that when people have higher incomes, they tend to prefer walking or driving alone to the rail station.

Another consistent trend, although in the opposite direction, is the decreasing number of access by bus as income increases. About half of those in the lowest three income categories (49.2%, 52.0%, and 48.3%) accessed the rail station by taking a bus first. In contrast, the only three income categories with less than 20% share taking the bus were also the three highest. Table 5 and Table 6 show that the higher income categories are largely underrepresented in riding the bus, while the lower income categories are largely overrepresented.

This may suggest that when populations have the resources to choose how to get the rail station, the preferred mode of access is to walk or to drive alone, compared to taking the bus. This also may indicate that a self-selection of transit ridership has occurred, in which those that choose to take the rail system (choice riders) live close enough to walk or have easy access via automobile. A greater number of riders in lower income categories may take transit out of necessity (transit dependent population) and may not live close enough to walk or have easy access to automobiles. This may also indicate a lack of affordable housing options that are close enough to MARTA rail for riders in this category to walk instead of taking the bus.

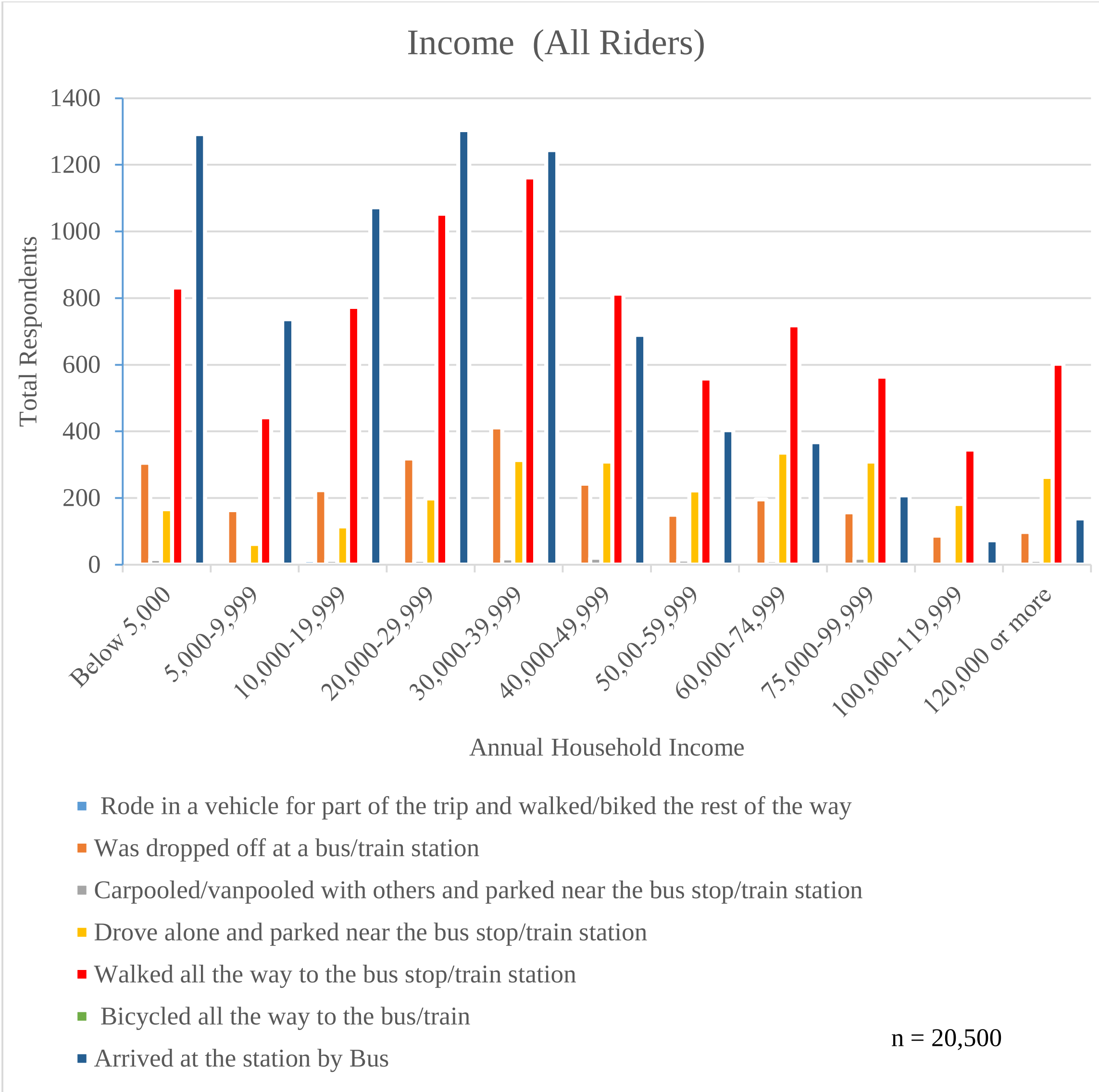

Figure 10: Income by mode of entry to the MARTA rail stations.

The problem with drawing conclusions around preferences in rail access with the ARC survey is that the populations may not be evenly distributed, and therefore may show that a certain group has a greater propensity to walk to transit, when they simply make up a larger proportion of the total population within walking distance. To understand the makeup of the population within walking distance to transit, American

Community Survey (ACS) data were downloaded and analyzed in Environmental Systems Research Institute's (ESRI's) ArcMap 10.3 and compared to the survey results for income, sex, age, ethnicity, and race. The finest geographic detail available for household income groups was at the block group level. To compare the ACS data to the ARC survey data, both datasets needed to be grouped to create common income categories. This yielded a total of nine categories all shown in Figure 11: Less than $10,000; $10,000 to $19,999; $20,000 to $29,999; $30,000 to $39,999; $40,000 to $49,999; $50,000 to $59,999; $60,000 to $74,999; $75,000 to $99,999; and $100,000 or More.

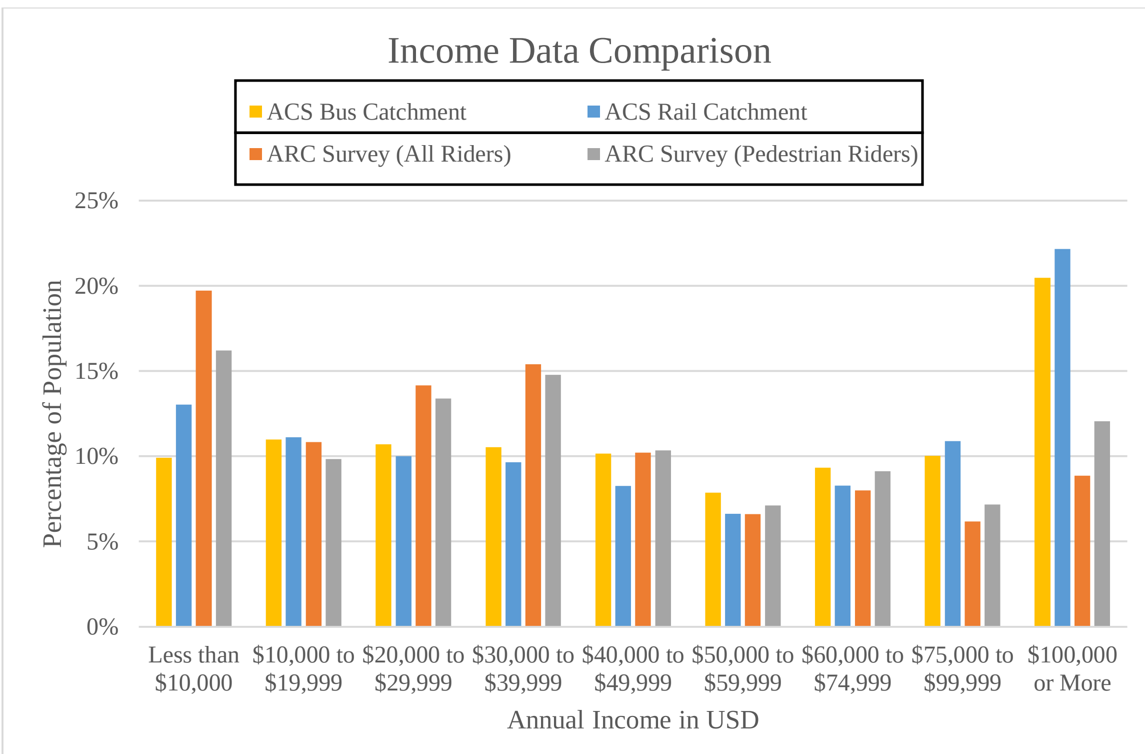

Figure 11: Comparison of household income composition between ACS data and the ARC survey data.

In total, there is an estimated 47,453 total households within a 0.5 mile buffer of the 38 MARTA stations using the ACS data. The most common category for the entire system was also the wealthiest, which was $100,000 or more with 10,514 households (or 22.16% of the total). The next most common was the least wealthy category of less than $10,000, with 6,185 households (or 13.0% of the total). Overall, there is a pattern showing the highest and lowest income categories being the most common, with the middle categories being the least common, with only 3,144 (6.6%) in the $50,000 to $59,999 category.

Several of the income categories contrast with the results from the survey. First, the survey resulted with a much smaller percentage of MARTA rail riders from the highest income category. A lower proportion of the highest income category for transit riders compared to census data would suggest that in general, high income households are not as likely to take MARTA rail as other income categories. The MARTA rail system appears to capture a smaller percentage of the riders of high income households. Interestingly, the ACS rail catchment has a greater percentage of people in the higher income categories than in the bus catchment, suggesting that more higher incomes households are clustered around rail stations rather than bus stops. This may be due to increases in property values around rail stations, which would also help explain the increased number of higher incomes that walk to the rail station rather than take the bus to get to the rail station.

Another difference in survey data compared to the ACS catchment data is that a greater percentage of MARTA riders are in the lower income categories than residents in the ACS catchments. This corroborates pervious research that states that lower incomes populations are more likely to take transit than higher income individuals. Overall, the comparison of the ACS data to the ARC survey data suggest that a greater percentage of lower income households are taking MARTA rail. MARTA rail appears to be capturing a greater percentage of the lower income market than other markets. The comparison also suggests that the highest income categories have a greater percentage of residents living closer to rail, and subsequently walking to access rail stations.

The income distribution varies widely at the station level. Using the ARC survey for all riders, there were five stations in the system with more than 25% of the households with less than $10,000 in income: West End with 29.8%, Oakland City with 28.8%, Peachtree Center with 26.7%, Bankhead with 26.4%, and Vine City with 26.3%. Some of the lowest percentage of riders with less than $10,000 in income occurred at Medical Center with 2.1%, East Lake with 3.3%, Indian Creek with 3.9%, and Dunwoody with 4.17%. Across all data methods, a trend of lower income riders/residents tend to occur along the south and west segments of the rail alignment, and higher income riders/residents tend to occur along the east and north segments of the rail alignment.

In light of the ARC survey and ACS income analysis, the inclusion of an income category in the DRM should benefit the ridership estimates. Overall the statistics suggest that lower income individuals are more likely to take transit than higher income individuals. However, the lower income groups are underrepresented in the pedestrian access mode compared to the higher income groups. This may reveal a lack of lower income housing supply within walking distance to MARTA rail stations. MARTA therefore may benefit by building affordable housing options near MARTA stations to capitalize on ridership potential in the new developments. With the disproportionate number of high income riders preferring to walk or drive alone to rail stations, inclusion of housing options at multiple price points could produce a favorable and balanced approach to housing near MARTA stations.

4.1.4 Age

Another piece of information requested via the ARC survey concerned age, shown graphically in Figure 12 with a total of 21,304 survey responses. The most common response was 18-24 (with 5,856), followed by 25-34 (with 5,418), 35-44 (with 3,993) and 45-54 (with 3,292). The categories "65 of more" and "Under 18" had the fewest numbers of respondents, with 421 and 514 respondents each, respectively.

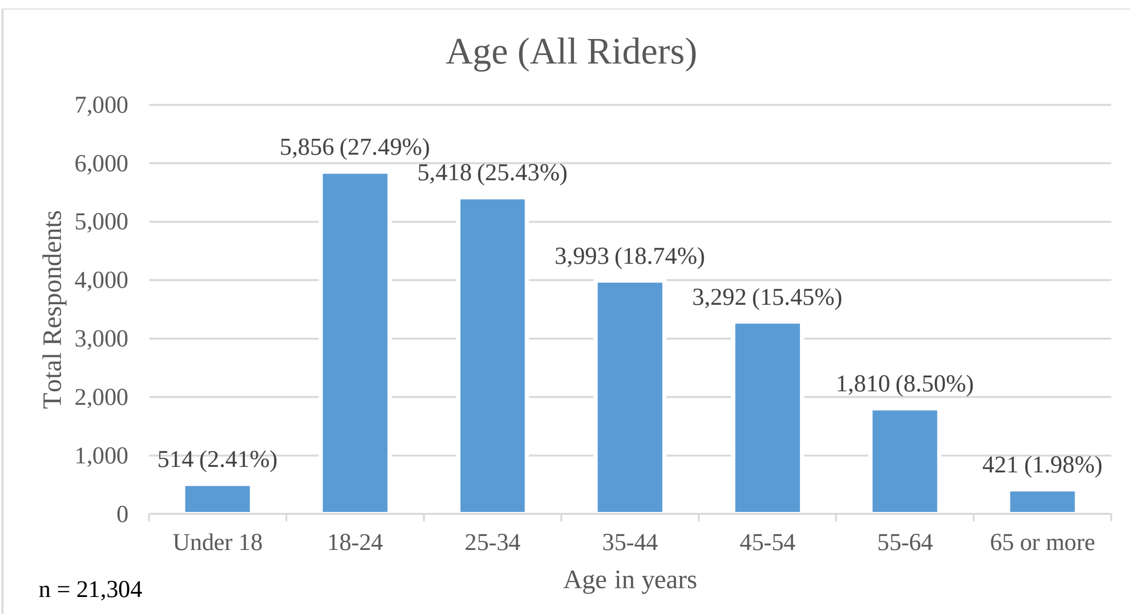

Figure 12: ARC survey results of age of those entering the MARTA system.

To help understand the potential relationship between age and walking to transit, among other modes, contingency tables were calculated in Table 8 and the chi-squared statistic is shown in Table 9. The chi-squared statistic of 352.4 is larger than the five percent critical value of 50.7, which indicates the rejection of the null hypothesis that age and walking to transit are independent. Inspecting the cross tabulated data of observed and expected counts reveals that there are several large disparities that may indicate a relationship between the two variables.

Table 8: Cross tabulated mode of access and age.

			Age							
			under 18	18-24	25-34	35-44	45-54	55-64	65 or more	Total
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	0.0	16.0	19.0	16.0	6.0	7.0	1.0	65.0
		Expected Count	1.6	17.9	16.5	12.2	10.0	5.5	1.3	65.0
	Was dropped off at a bus/train station	Count	68.0	789.0	662.0	428.0	284.0	180.0	41.0	2,452.0
		Expected Count	59.2	674.0	623.6	459.6	378.9	208.3	48.5	2,452.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	2.0	52.0	44.0	30.0	21.0	15.0	4.0	168.0
		Expected Count	4.1	46.2	42.7	31.5	26.0	14.3	3.3	168.0
	Drove alone and parked near the bus stop/train station	Count	12.0	521.0	626.0	472.0	534.0	316.0	66.0	2,547.0
		Expected Count	61.5	700.1	647.7	477.4	393.6	216.4	50.3	2,547.0
	Walked all the way to the bus stop/train station	Count	152.0	2,212.0	2,208.0	1,605.0	1,236.0	689.0	140.0	8,242.0
		Expected Count	198.9	2,265.5	2,096.1	1,544.8	1,273.6	700.2	162.9	8,242.0
	Bicycled all the way to the bus/train	Count	0.0	15.0	18.0	9.0	6.0	0.0	0.0	48.0
		Expected Count	1.2	13.2	12.2	9.0	7.4	4.1	0.9	48.0
	Rode the bus to the station	Count	280.0	2,251.0	1,841.0	1,433.0	1,205.0	603.0	169.0	7,782.0
		Expected Count	187.8	2,139.1	1,979.1	1,458.6	1,202.5	661.2	153.8	7,782.0
Total		Count	514.0	5,856.0	5,418.0	3,993.0	3,292.0	1,810.0	421.0	21,304.0
		Expected Count	514.0	5,856.0	5,418.0	3,993.0	3,292.0	1,810.0	421.0	21,304.0

Table 9: Chi-Squared Test for Independence for the mode of access and income.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	352.368[a]	36	0.000
Likelihood Ratio	370.036	36	0.000
N of Valid Cases	21,304		

a. 7 cells (14.3%) have expected count less than 5. (the minimum expected count is 0.95)

H0: The mode of MARTA rail access and age are independent.
Ha: The mode of MARTA rail access and age are not independent.

While the contingency tables do not show the large disparities in demographic and mode choice seen in the income variable, subtle patterns are still identified. Driving alone is observed less than expected in the age groups 34 and under, while driving alone is observed more than expected in the age groups 45 and over (ages 35 through 44 have negligible differences in the observed and expected counts). This may suggest that older transit riders tend to drive alone to get to transit more often that their younger counterparts. The contingency tables also show that the middle aged categories (25 through 34 and 35 through 44) have slightly higher observed walking counts than expected, while the younger and older age categories have slightly fewer observed counts than expected counts. This may suggest that middle aged transit riders prefer walking more often than younger or older transit riders, or that an interaction with another variable is occurring, such as income, housing choice, etc. It appears that with transit riders in the youngest two age categories being underrepresented in the driving alone and walking mode choices, they instead choose to ride the bus to the stations, shown with a greater observed than expected count in Table 8.

Additional support in graphical form is shown in Figure 13 below. Overall, Figure 13 shows that the youngest age category (under 18) had the lowest percentage of people walking to the rail station with only 29.57%. The under 18 category also had the greatest percentage of those taking the bus to get to the station, with 54.47%. Interestingly, the oldest age category of 65 and over had a similar access mode share, with the only 33.25% walking to the station and a relatively high percentage of those taking the bus to the rail station with 40.14%. The age categories that had the greatest percentage of people walking to the rail station were the 25 to 34 category and the 35 to 44 category, both of which had greater than 40% of the transit users accessing on foot.

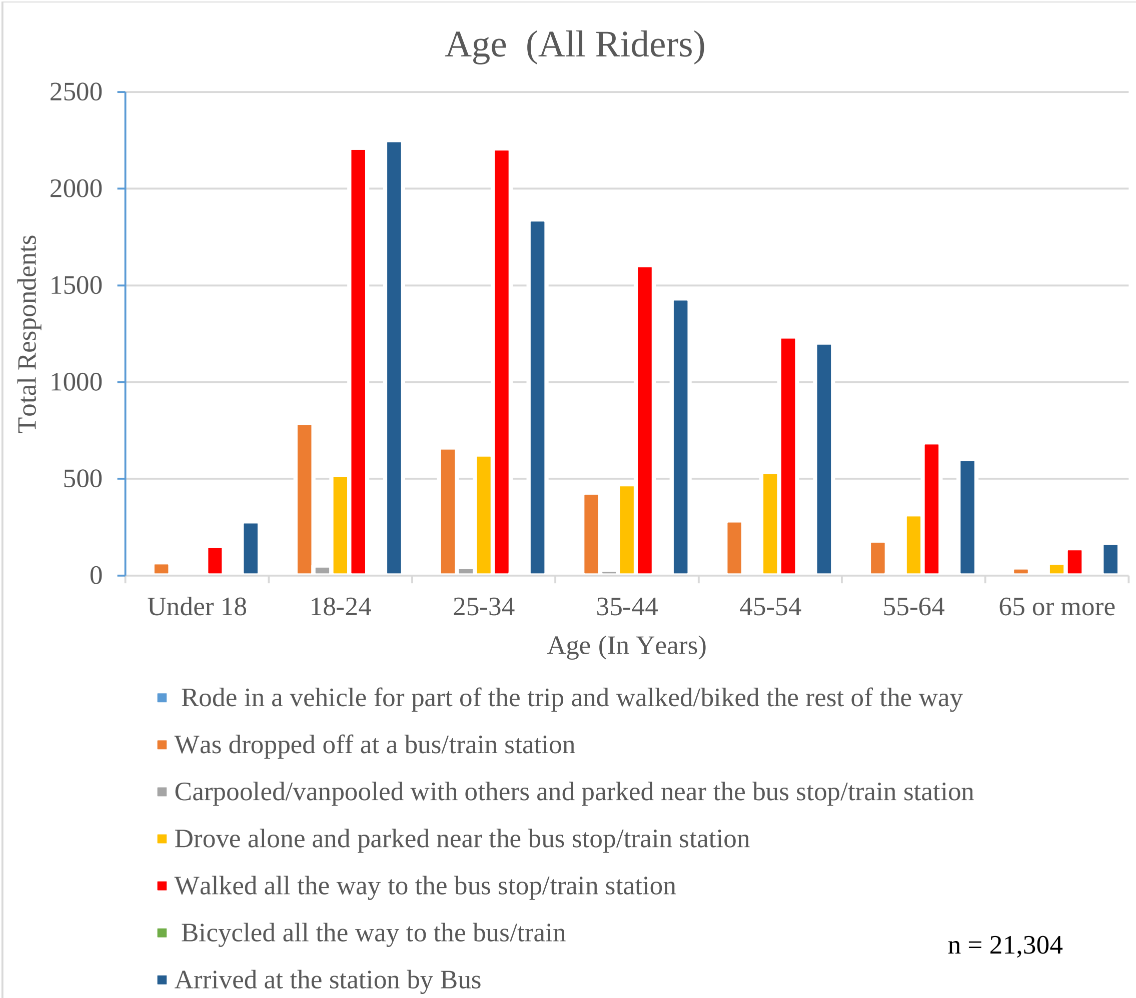

Figure 13: Age by mode of entry to the MARTA rail stations.

Figure 14 below provides a comparison of the ARC survey data and census data from the ACS. A general sense of how well MARTA is capturing the different age markets can be gleaned from comparing the distribution of age in the ACS data to the ARC survey data. In general, the under 18 and over 65 age categories have the greatest disparity in distribution between the two datasets. This suggests that the youngest and oldest age groups in the rail and bus catchments taking transit at a lower rate than the middle age groups. It also appears that a greater percentage people in the 18 to 24 and 25 to 34 age categories live close to rail, while all other age categories have a greater percentage of people living in bus catchment areas. One reason that the 18 to 34 age group could be living closer to rail stations is for quick transit access to jobs. In addition, the same two age groups (18 to 24 and 25 to 34) are taking rail transit than other groups.

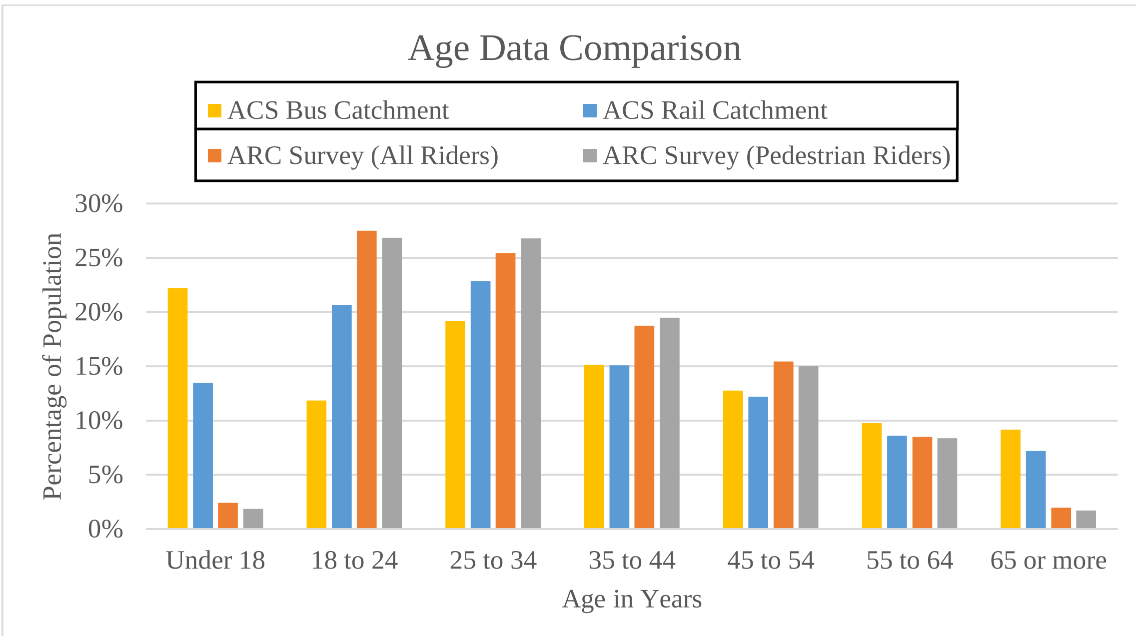

Figure 14: Comparison of age composition between ACS data and the ARC survey data.

For the purposes of constructing a DRM that relates rail transit catchments to ridership, special attention should be paid to the age groups of 18 to 24, 25 to 34, and 35 to 44. With evidence from the ARC survey and the ACS census data, these groups tend to take transit at a greater rate than other age groups. Moreover, these age groups are more likely to walk to transit than their older or younger neighbors.

4.1.5 Gender

Out of the 21,302 rail riders who answered the survey question on gender, 51.09% responded female and 48.91% responded male, shown graphically in Figure 15. However, of those who walked to the rail station, 52.52% were male and 47.48% were female, indicating that males were slightly more likely to walk to the rail stations than females.

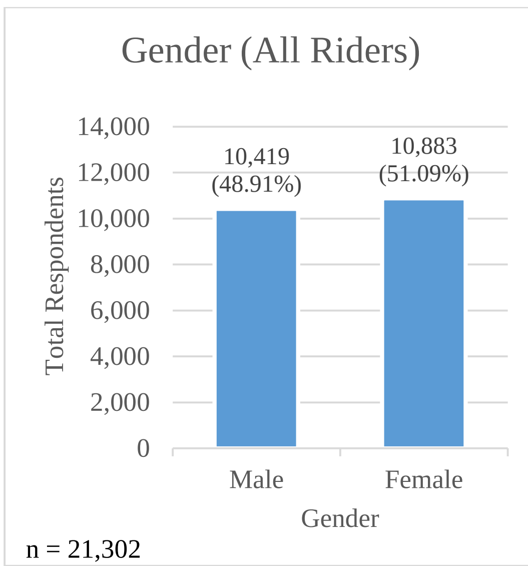

Figure 15: ARC survey results of gender of those entering the MARTA rail system.

Shown in Figure 16, the most common mode of accessing the rail station for males was walking (41.55%), followed by taking the bus (34.15%), being dropped off (11.64%), and driving alone (11.26%). For females, the most common mode of accessing rail stations was taking the bus (38.80%), followed by walking (35.96%), driving alone (12.63%), and being dropped off (11.38%). Overall, survey results show that males are more likely to walk to rail stations than females, who are slightly more likely to take a bus to a rail station. At the station level, 18 of the stations had more female respondents while 20 of the stations had more male respondents.

At the station level, the greatest disparity in gender occurred at Oakland City, where 61.99% were female and 38.01% were male. Other stations with a large percentage of female respondents were Medical Center (60.99% female and 39.01% male), Indian Creek (60.20% female and 39.80% male) and Kensington (59.79% female and 40.21% male). Conversely, stations with a large male presence were Garnett (60.45% male and 39.55% female), Sandy Springs (59.93% male and 40.07% female), Edgewood/Candler Park (58.86% male and 41.14% female), and Midtown (58.35% male and 41.65% female).

Figure 16: Gender by mode of entry to the MARTA rail stations.

As in the previous sections, cross tabulations and chi-squared statistics were calculated to gain a greater understanding of the relationship of gender and mode of access to the MARTA rail stations. The observed and expected counts in the contingency table are shown in Table 10, and the Chi-Squared Test for Independence is shown in Table 11. The chi-squared statistic of 100.0 is greater than the five percent critical value of 78.8. The greater chi-squared statistic indicates that the null hypothesis that age and mode of access to rail are independent can be rejected.

The cross-tabulations show that the observed count for males in the driving alone and riding the bus are less than the expected counts. The reverse is true for females, who drove alone and rode the bus more than the expected. Conversely, the observed count for males walking to the rail station was higher than the expected counts, while for females the observed count was lower than the expected counts. In general, the contingency tables show a slight male preference of walking instead of riding the bus or driving alone, and a slight female preference to drive alone or take the bus rather than walking.

Table 10: Cross tabulation of mode of access and gender.

			Gender		Total
			Male	Female	
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	34.0	31.0	65.0
		Expected Count	31.8	33.2	65.0
	Was dropped off at a bus/train station	Count	1,213.0	1,238.0	2,451.0
		Expected Count	1,199.8	1,252.6	2,451.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	75.0	93.0	168.0
		Expected Count	82.2	85.8	168.0
	Drove alone and parked near the bus stop/train station	Count	1,173.0	1,374.0	2,547.0
		Expected Count	1,245.8	1,301.2	2,547.0
	Walked all the way to the bus stop/train station	Count	4,329.0	3,913.0	8,242.0
		Expected Count	4,031.2	4,210.8	8,242.0
	Bicycled all the way to the bus/train	Count	37.0	11.0	48.0
		Expected Count	23.5	24.5	48.0
	Rode the bus to the station	Count	3,558.0	4,223.0	7,781.0
		Expected Count	3,805.8	3,975.2	7,781.0
Total		Count	10,419.0	10,883.0	21,302.0
		Expected Count	10,419.0	10,883.0	21,302.0

Table 11: Chi-Squared Test for Independence for the mode of access to rail stations and gender.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	100.039[a]	6	0.000
Likelihood Ratio	100.874	6	0.000
N of Valid Cases	21,302		

a. 0 cells (0.00%) have expected count less than 5. (the minimum expected count is 23.48)

H0: The mode of MARTA rail access and gender are independent.
Ha: The mode of MARTA rail access and gender are not independent.

ACS data were calculated around the bus stops and rail stations and compared to ARC survey data, which Figure 17 shows graphically. A greater percentage of females live in the bus catchments, while the reverse is true for rail catchments. The ARC survey data show the same trend, with more females taking rail transit than males, but more males walking to transit than females. These data suggest that males tend to live closer to rail stations and subsequently walk more to transit stations. There are more females living within walking distances to bus stops, which results in a greater number of females taking buses to rail stations. It is important, however, to note that even though the sample sizes for gender is large for each of these datasets, the relative differences between males and females are small.

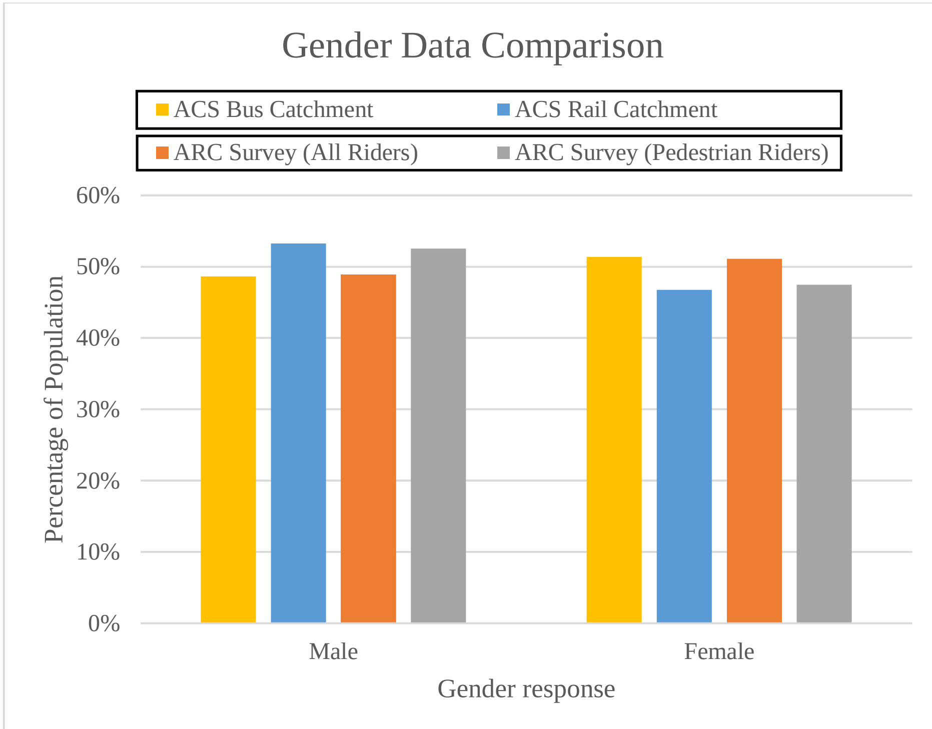

Figure 17: Composition of gender composition of ACS data and the ARC survey data.

4.1.6 Ethnicity

Of those surveyed, there were a total of 1,209 (5.68%) who answered “Yes” to the question “Are you Hispanic/Latino?”. The remaining 20,093 (94.32%) form the survey responded “No”. Figure 18 shows the total number of each category graphically.

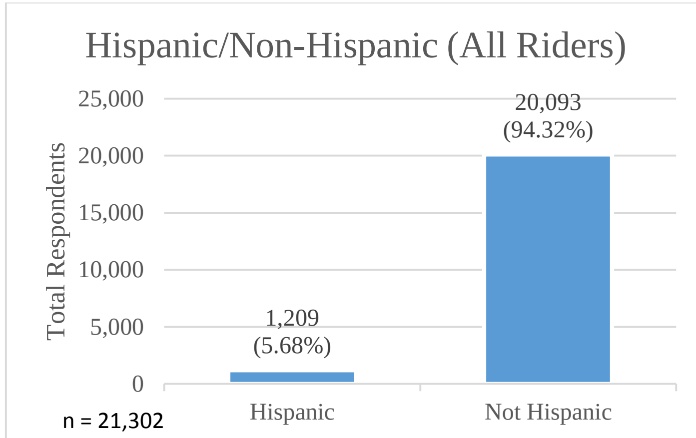

Figure 18: ARC survey of ethnicity of those entering the MARTA rail system.

A deeper investigation of Hispanic/Non-Hispanic riders and mode of access to the rail stations is shown in Figure 19. Of transit riders that indicated Hispanic ethnicity, the most common mode to the rail stations was to walk with 46.73%, followed by taking the bus (27.96%), driving alone (12.90%), and being dropped off (11.08%). Of those who answered “No” to the same question, the most common was still walking, but only at 38.21%. This was followed by taking the bus (37.04%), driving alone (11.90%), and being dropped off (11.53%). This suggests that having Hispanic ethnicity increases the likelihood of walking to MARTA rail. All other modes were distributed similarly to the Non-Hispanic riders, except for riding the bus. The ARC survey results show that the apparent increase in those walking to the rail station coincides with a decrease in those taking the bus to rail stations.

At the station level, the greatest percentage of respondents that answered “Yes” to this question occurred at Chamblee with (13.70%), Lindbergh (12.31%), Doraville (12.23%), Brookhaven (10.00%), and Peachtree Center (9.51%). The top four stations in terms of percentage of Hispanics/Latinos occur all on the Gold Line, located on the northeastern section of the MARTA rail system. The lowest percentages of those answering "Yes" to this question occurred at West Lake (1.13%), Oakland City (1.50%), Vine City (1.54%) Decatur (2.19%) and West End (2.78%). Aside from Decatur, all of these stations are west or south of Five Points Center Station.

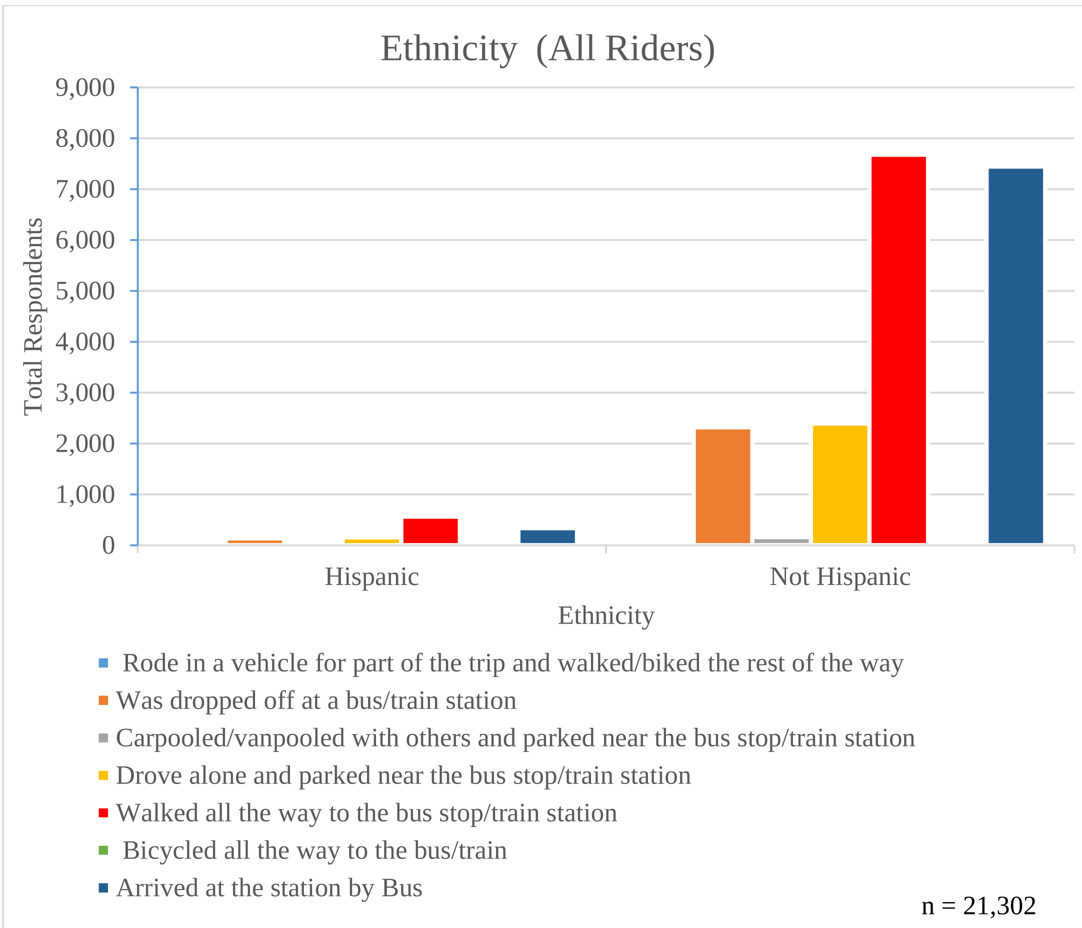

Figure 19: Ethnicity by mode of entry to the MARTA rail stations.

Contingency tables have been calculated by cross-classifying (or cross-tabulating) the data to see if the ethnicity variable differs by mode of rail station access. Table 12 shows the observed counts and the expected counts for the mode of access to the rail station and ethnicity, and Table 13 shows the Chi-Squared statistic for the test of independence. The chi-squared statistic was 49.8 and with a five percent critical value of 12.6, a rejection of the null hypothesis is justified. Similar to the previous section, the null hypothesis for this variable is that ethnicity and mode of MARTA rail access are independent.

Table 12: Cross tabulated mode of access and ethnicity.

			Hispanic/Not Hispanic		Total
			Hispanic Origin	Not Hispanic Origin	
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	4.0	61.0	65.0
		Expected Count	3.7	61.3	65.0
	Was dropped off at a bus/train station	Count	134.0	2,317.0	2,451.0
		Expected Count	139.1	2,311.9	2,451.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	11.0	157.0	168.0
		Expected Count	9.5	158.5	168.0
	Drove alone and parked near the bus stop/train station	Count	156.0	2,391.0	2,547.0
		Expected Count	144.6	2,402.4	2,547.0
	Walked all the way to the bus stop/train station	Count	565.0	7,677.0	8,242.0
		Expected Count	467.8	7,774.2	8,242.0
	Bicycled all the way to the bus/train	Count	1.0	47.0	48.0
		Expected Count	2.7	45.3	48.0
	Rode the bus to the station	Count	338.0	7,443.0	7,781.0
		Expected Count	441.6	7,339.4	7,781.0
Total		Count	1,209.0	20,093.0	21,302.0
		Expected Count	1,209.0	20,093.0	21,302.0

Table 13: Chi-Squared Test for Independence for the mode of access to rail stations and ethnicity.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	49.778[a]	6	0.000
Likelihood Ratio	50.988	6	0.000
N of Valid Cases	21,302		

a. 2 cells (14.3%) have expected count less than 5.
(the minimum expected count is 2.72)

H0: The mode of MARTA rail access and ethnicity are independent.
Ha: The mode of MARTA rail access and ethnicity are not independent.

The cross tabulations show similar trends revealed from Figure 19. For riders with Hispanic origin, the observed count for walking to the rail station is greater than the expected count. Conversely, riders without Hispanic origin have a lower observed count than expected count, indicating that those with Hispanic origin use transit at a greater rate than those without Hispanic origin. Although not as strong, driving to the station alone had the same trend, with riders that had Hispanic origin recording more observed counts than expected and those without Hispanic origin recording fewer observed counts than expected. Another disparity in observed versus expected occurs with riding the bus to the station, where riders with Hispanic origin had fewer observed counts than expected and riders without Hispanic origin had greater observed counts than expected counts.

Figure 20 compares the ACS census data around transit catchments to the ARC survey data on Hispanic origin. The ACS data show that the Hispanic origin population has a greater percentage of the population around bus stops than rail stations. Despite this trend, the pedestrian riders have a larger percent with Hispanic origin than for other modes of access to rail. In general, the data sources are indicating that the Hispanic origin population have a greater preference for walking compared to those that are not of Hispanic origin. Despite this preference, the Hispanic origin population is more concentrated around bus stops than rail stations. The counterintuitive results shown by the survey and census comparison of Hispanic population may make it a less impactful variable in the DRM, but the propensity to walk to rail transit by the Hispanic population provides enough evidence to test it in the model.

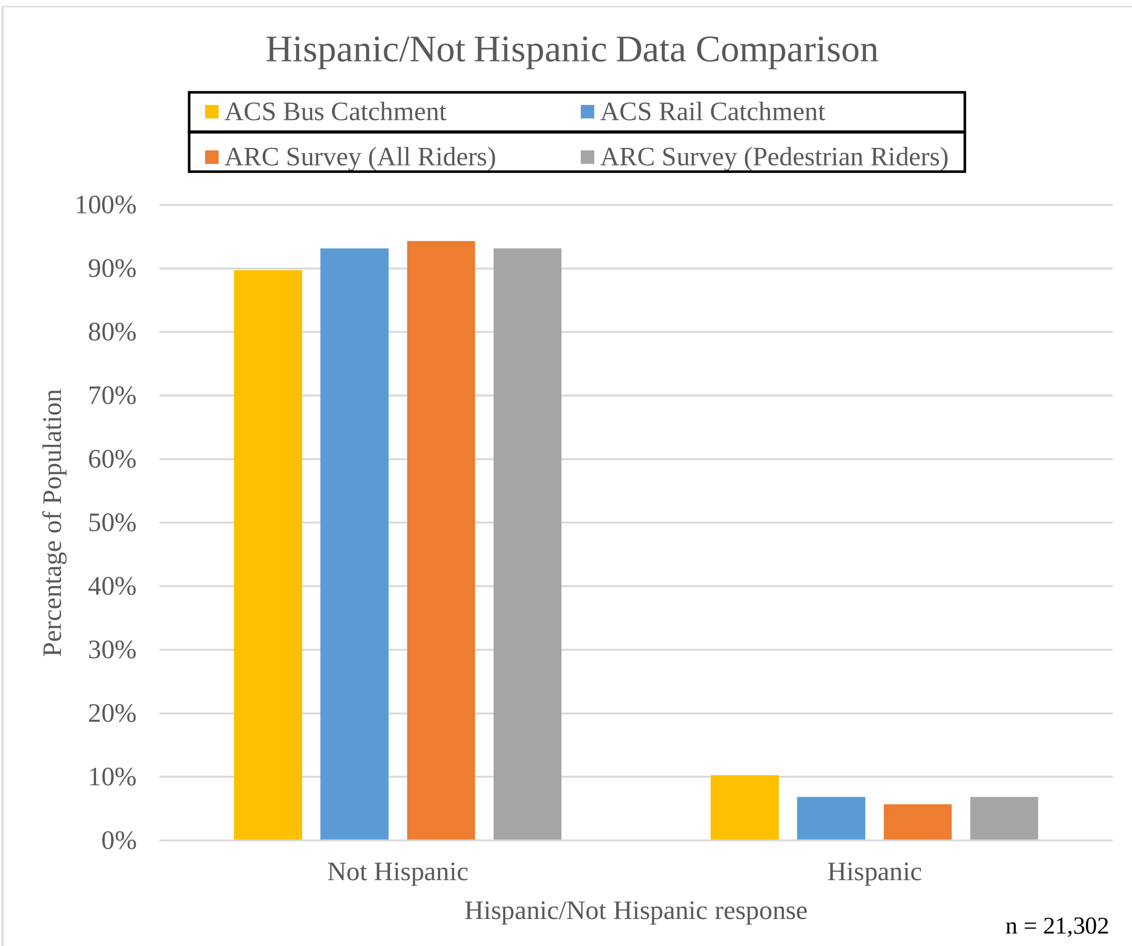

Figure 20: Comparison of ethnicity between ACS data and the ARC survey data.

4.1.7 Race

The ARC survey showed that transit ridership among different racial groups is highly uneven. When asked "How would you describe your race?", 68.5% responded Black/African American, 23.3% responded White, 4.7% responded Other, 2.5% responded Asian, and 1.1% responded American Indian. Results are shown graphically in Figure 21 below.

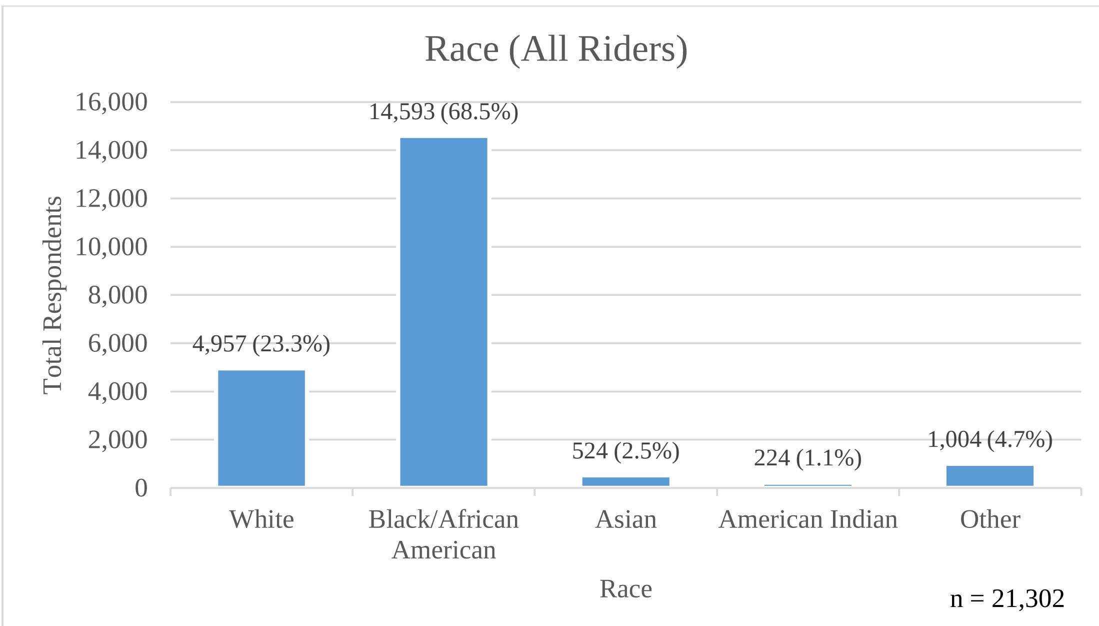

Figure 21: ARC survey results of race of those entering the MARTA rail system.

Cross tabulations of the observed counts versus the expected counts were calculated and shown below in Table 14. In addition, the Chi-Squared Test for Independence results are shown in Table 15, with a null hypothesis stating that race and mode of MARTA rail access are independent. With a chi-squared statistic of 2,131.9 and a five percent critical value of 36.4, it is justifiable to reject the null hypothesis for this variable as well.

Table 14: Cross tabulated mode of access and race.

			Race/Ethnicity					
			White	Black/African American	Asian	American Indian	Other	Total
Mode	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Count	19.0	41.0	4.0	0.0	1.0	65.0
		Expected Count	15.1	44.5	1.6	0.7	3.1	65.0
	Was dropped off at a bus/train station	Count	566.0	1,687.0	49.0	33.0	116.0	2,451.0
		Expected Count	570.4	1,679.1	60.3	25.8	115.5	2,451.0
	Carpooled/vanpooled with others and parked near the bus stop/train station	Count	75.0	72.0	13.0	0.0	8.0	168.0
		Expected Count	39.1	115.1	4.1	1.8	7.9	168.0
	Drove alone and parked near the bus stop/train station	Count	1,010.0	1,300.0	100.0	23.0	114.0	2,547.0
		Expected Count	592.7	1,744.8	62.7	26.8	120.0	2,547.0
	Walked all the way to the bus stop/train station	Count	2,567.0	4,781.0	283.0	132.0	479.0	8,242.0
		Expected Count	1,917.9	5,646.2	202.7	86.7	388.5	8,242.0
	Bicycled all the way to the bus/train	Count	27.0	17.0	3.0	0.0	1.0	48.0
		Expected Count	11.2	32.9	1.2	0.5	2.3	48.0
	Rode the bus to the station	Count	693.0	6,695.0	72.0	36.0	285.0	7,781.0
		Expected Count	1,810.6	5,330.4	191.4	81.8	366.7	7,781.0
Total		Count	4,957.0	14,593.0	524.0	224.0	1,004.0	21,302.0
		Expected Count	4,957.0	14,593.0	524.0	224.0	1,004.0	21,302.0

Table 15: Chi-Squared Test for Independence for the mode of access and race.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2131.943[a]	24	0.000
Likelihood Ratio	2270.721	24	0.000
N of Valid Cases	21,302		

a. 8 cells (22.9%) have expected count less than 5. (the minimum expected count is .50)

H0: The mode of MARTA rail access and race are independent.

Ha: The mode of MARTA rail access and race are not independent.

Overall, those who walked to access the rail stations were most likely to be Black/African American (58.01% percent of the total). This was followed by White (31.15%), Other (5.81%), Asian (3.43%), and American Indian (1.60%). Figure 22 shows the uneven distribution of mode of rail access by race. Those who identified as Black/African American were most likely to take the bus to get the rail station, with (45.88%), followed by walking (32.76%) while driving alone made up only 8.91% of the Black/African American transit rider access modes. This contrasted with the mode choice of the White transit riders, of whom 51.79% walked, 13.98% took the bus, and 20.38% drove alone. The mode of access percentages for Asian transit riders were very similar to White transit riders, with 54.01% walking, 13.74% taking the bus, and 19.08% driving alone. Walking to transit was the most common mode of access for American Indians with 58.93%. After walking to rail, American Indians were likely to ride the bus (16.07%), followed by being dropped off (14.73%), and drive alone (10.27%). Those who indicated "Other" for their race, were most likely to walk to rail stations with 47.71%, followed by taking the bus (28.39%), getting dropped off (11.55%), and driving alone (11.35%). The ARC survey shows that although Black/African American transit riders make up the most patrons who walk to transit, taking the bus is a more likely alternative for this racial group. The remaining four racial groups, all of whom get to the rail stations by walk access more than other access modes, make up the remaining 41.99% of those who walked to rail stations.

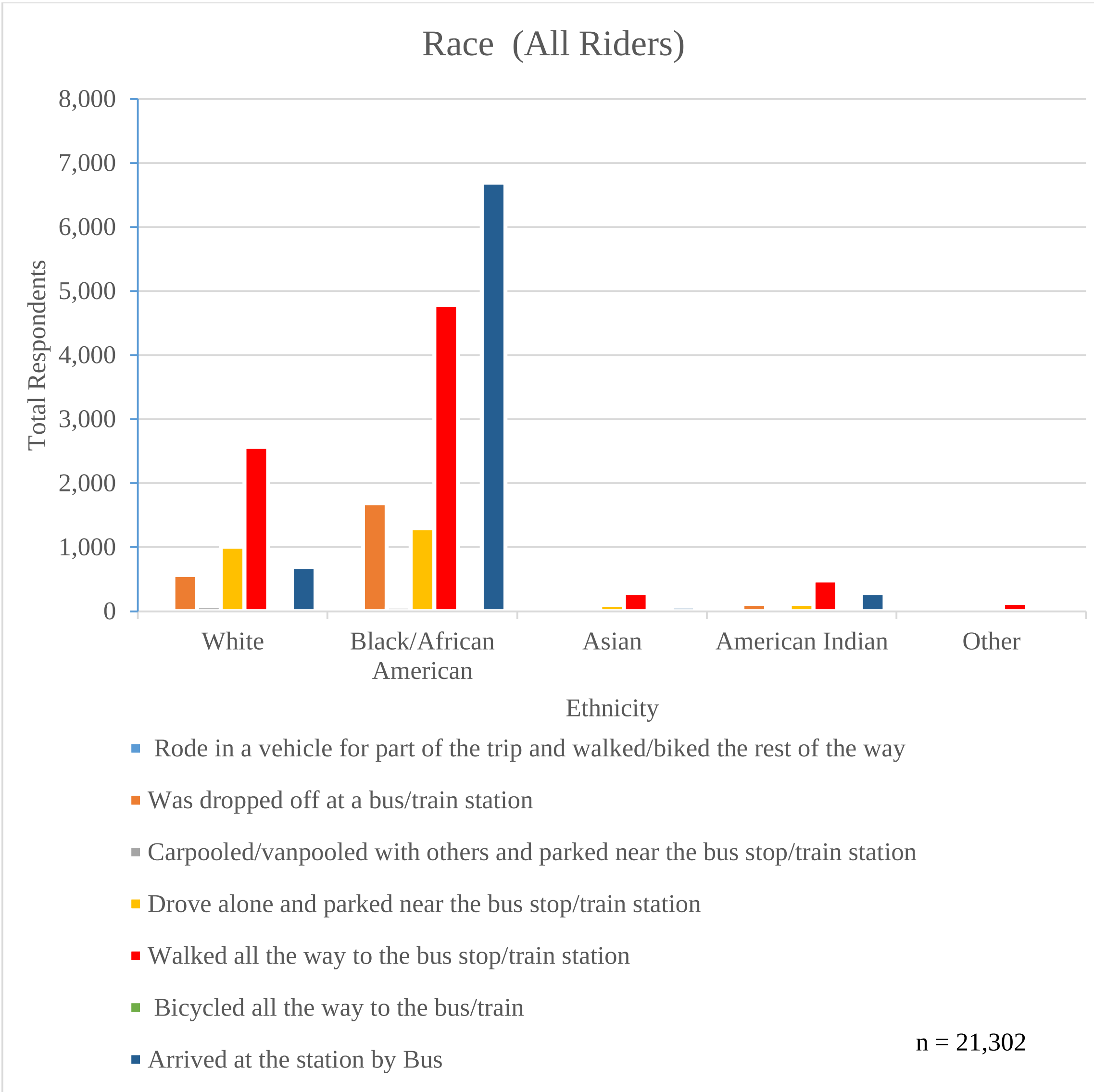

Figure 22: Race by mode of entry to the MARTA rail stations.

The ARC survey showed that racial composition of populations accessing the rail stations varies greatly. At the station level, numerous stations had 90% or more Black/African American respondents, including West Lake (96.24%), West End (94.99%), Bankhead (93.98%), Vine City (93.85%), Oakland City (93.82%), Ashby (93.20%), and Hamilton E Holmes (90.21%). All of these stations except for West End are along the western section of the Blue/Green rail lines. The greatest percentage of White respondents occurred at North Springs with 51.24%, followed by 45.72% at Peachtree Center, 42.86% at Medical Center, 41.77% at Edgewood/Candler Park, and 40.71% at Brookhaven/Oglethorpe. Relatively large percentages of the response "Other" were recorded for Lindbergh (10.70%), Chamblee (10.08%), Lindbergh (10.70%), Midtown (7.38%), and Peachtree Center (7.22%). Only six stations had more than 5% Asian respondents. These included Sandy Springs (9.06%), Doraville (7.10%), Buckhead (6.27%), Brookhaven (6.07%), Peachtree Center (5.89%), and Midtown (5.10%). The least common response for race was American Indian, which had the greatest percentages at Lenox (3.38%), Midtown (2.99%), North Springs (2.94%), Peachtree Center (2.85%), and North Avenue (2.03%).

Figure 23 shows the comparison of the ARC survey data to the census demographic data surrounding bus stops and rail stations. Of those that live in rail catchments, a larger percentage of percentage from census data are White and Asian compared to the population that live in bus catchments. Additionally, the ARC survey shows the same relationship, where the population getting to rail stations via walk access had a greater percentage of White and Asian respondents than the population getting to rail by any mode. Conversely, the Black population showed the opposite trend, where a greater percentage of the population living within the bus catchment was Black compared to the population living within the rail catchment. Moreover, when observing the ARC survey data, the population that arrived to the station by foot were made up of a greater percentage of Black population compared to the survey data of all riders. While these results show that the Black population has a greater propensity to arrive to the station by bus than by foot, the overall number that the Black population contributes to taking transit is much greater than any other racial group. White the ACS rail catchment shows that similar numbers of Black and White populations live near rail stations, the ARC survey shows that the Black population is taking rail at a much greater rate, both in the all mode category as well as in the pedestrian rider category.

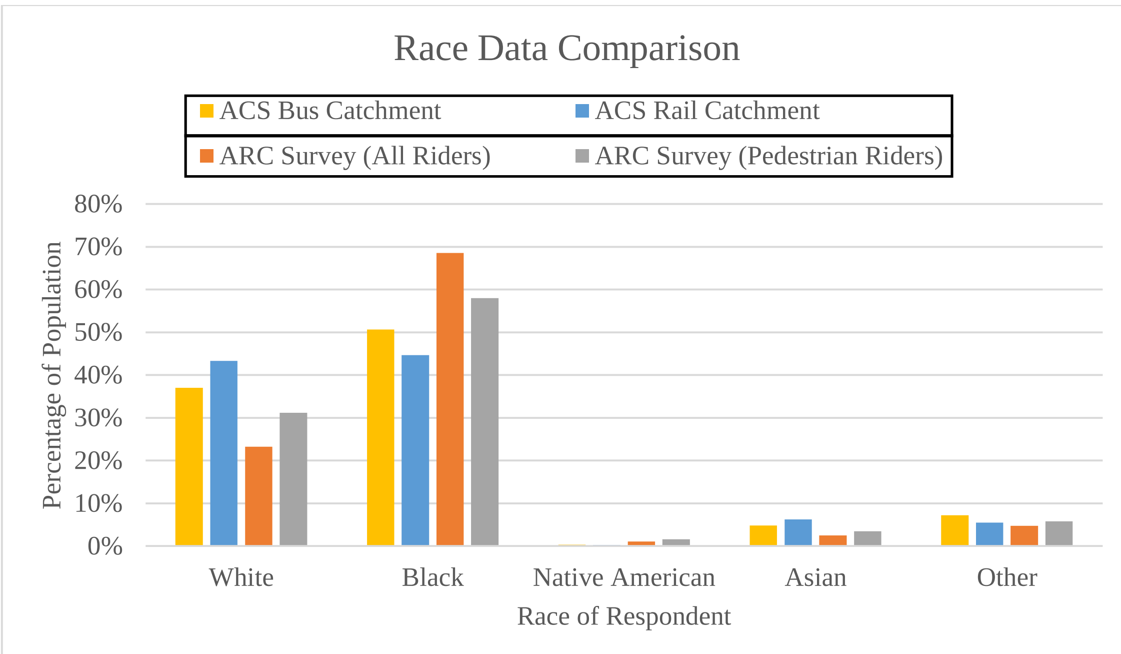

Figure 23: Comparison of race composition of ACS data and the ARC survey data.

For the purposes of constructing a DRM, the ARC survey and the ACS data show that race is an important factor in contributing to transit ridership, both as a whole and by pedestrian access. Inclusion of a race variable has the potential to increase the predictive power of the ridership model for pedestrian access ridership in the Atlanta metropolitan area.

4.2 Environmental and station characteristics variable results

Results of the environmental variables introduced in Chapter 3 are revealed here, which include the land use mix index (LUMI), street density, intersection density, and parking. The LUMI indicates land use the heterogeneity and employs the square footage of commercial, residential, and industrial land-use codes at each of the stations. The LUMI values can have a low of 0 (meaning completely homogenous land uses) to a high of 1 (meaning perfect balance of land uses), shown in Table 16. The average of the LUMI for the stations was 0.558. A total of 17 of the 38 stations fell within a range of 0.40 – 0.70. The lowest value on the LUMI occurred at the Airport, with a score of 0.000, meaning only one type of land use was recorded in the transit catchment. The other stations with little land-use mix occurred at East Lake (0.105), Medical Center, (0.129), Peachtree Center (0.150), and Sandy Springs (0.212). The stations with the most balanced land uses were Chamblee (0.996), Bankhead (0.938), East Point (0.894), Doraville (0.881), and West End (0.838). Overall, the LUMI values varied without a clear geographic trend across the metropolitan region.

Table 16: Environmental and station characteristics variables by rail station

Station	LUMI	Street Density (miles per square mile)	Intersection Density (per square mile)	Parking Spaces
Five Points	0.30	28.3	862.6	0
GA State	0.58	25.6	628.7	0
King Memorial	0.82	21.7	494.3	0
Inman Park	0.54	20.4	502.5	366
Edgewood/Candler	0.37	18.5	254.0	679
East Lake	0.11	17.3	225.4	611
Decatur	0.69	19.6	387.2	0
Avondale	0.78	16.9	329.3	823
Kensington	0.62	11.0	115.9	1,946
Indian Creek	0.63	11.9	138.8	2,350
Peachtree Center	0.15	27.3	766.9	0
Sandy Springs	0.21	16.1	265.3	1,170
North Springs	0.55	10.7	156.6	2,325
Civic Center	0.53	27.8	824.2	0
North Avenue	0.36	24.5	692.7	0
Midtown	0.65	22.3	446.2	0
Art Center	0.52	22.8	534.9	0
Lindbergh	0.61	16.0	286.5	2,907
Buckhead	0.43	17.1	275.4	0
Medical Center	0.13	12.8	140.1	0
Dunwoody	0.30	9.2	93.3	1,048
Doraville	0.88	11.2	173.2	1,070
Lenox	0.54	14.9	201.6	321
Brookhaven	0.54	15.8	245.8	1,252
Chamblee	1.00	12.6	132.4	1,713
Bankhead	0.94	12.3	336.2	0
Garnett	0.78	31.7	881.9	0
West End	0.84	21.8	483.9	547
Oakland City	0.78	17.8	276.3	337
Lakewood/Fort McPherson	0.83	17.4	294.2	1,134
East Point	0.89	21.3	449.5	927
College Park	0.51	20.6	375.4	1,883
Airport	0.00	13.7	274.4	0
Dome/GWCC/Phillips/CNN	0.36	17.5	485.6	0
Vine City	0.76	22.3	662.0	27
Ashby	0.47	22.0	523.2	153
West Lake	0.57	12.4	168.1	391
Hamilton E Holmes	0.64	17.6	336.2	1,436

A cartographic representation of street density is shown in Figure 24, with densities around each of the MARTA rail stations in shades of blue (light blue indicating a less dense street network and dark blue indicating a denser street network). The MARTA stations with the greatest street density (in miles of street per square mile of catchment) were Garnett (31.7), Five Points (28.3), Civic Center (27.8), Peachtree Center (27.3), and GA State (25.6). The MARTA stations with the lowest densities were Dunwoody (9.2), North Springs (10.7), Kensington (11.0), Doraville (11.2), and Indian Creek (11.8). Overall, the stations close to the Downtown and Midtown locations observed the greatest densities, while the stations farthest from Downtown especially on the Red and Gold northern lines, had the lowest street density. Street density can act as a barrier to transit access in the case of disconnected street networks, or a support for transit access when blocks are short and out of direction travel is kept to a minimum.

Similar to street density, intersection density (number of intersections per square mile of catchment) can provide aid to pedestrian access by way of greater opportunities to change paths to arrive at the destination. Results from calculating the intersection density revealed similar conclusions to street density, also shown in Table 16. The stations of Garnett (881.9), Five Points (862.6), Civic Center (824.2), Peachtree Center (766.9), and North Avenue (692.7), had the greatest intersection densities. The stations of Dunwoody (93.3), Kensington (115.9), Chamblee (132.4), Indian Creek (138.9), and Medical Center (140.1), had the lowest intersection densities in the system. For intersection density, overall the Downtown and Midtown stations observed the greatest densities and the periphery and northern Red and Gold Line stations had the lowest densities.

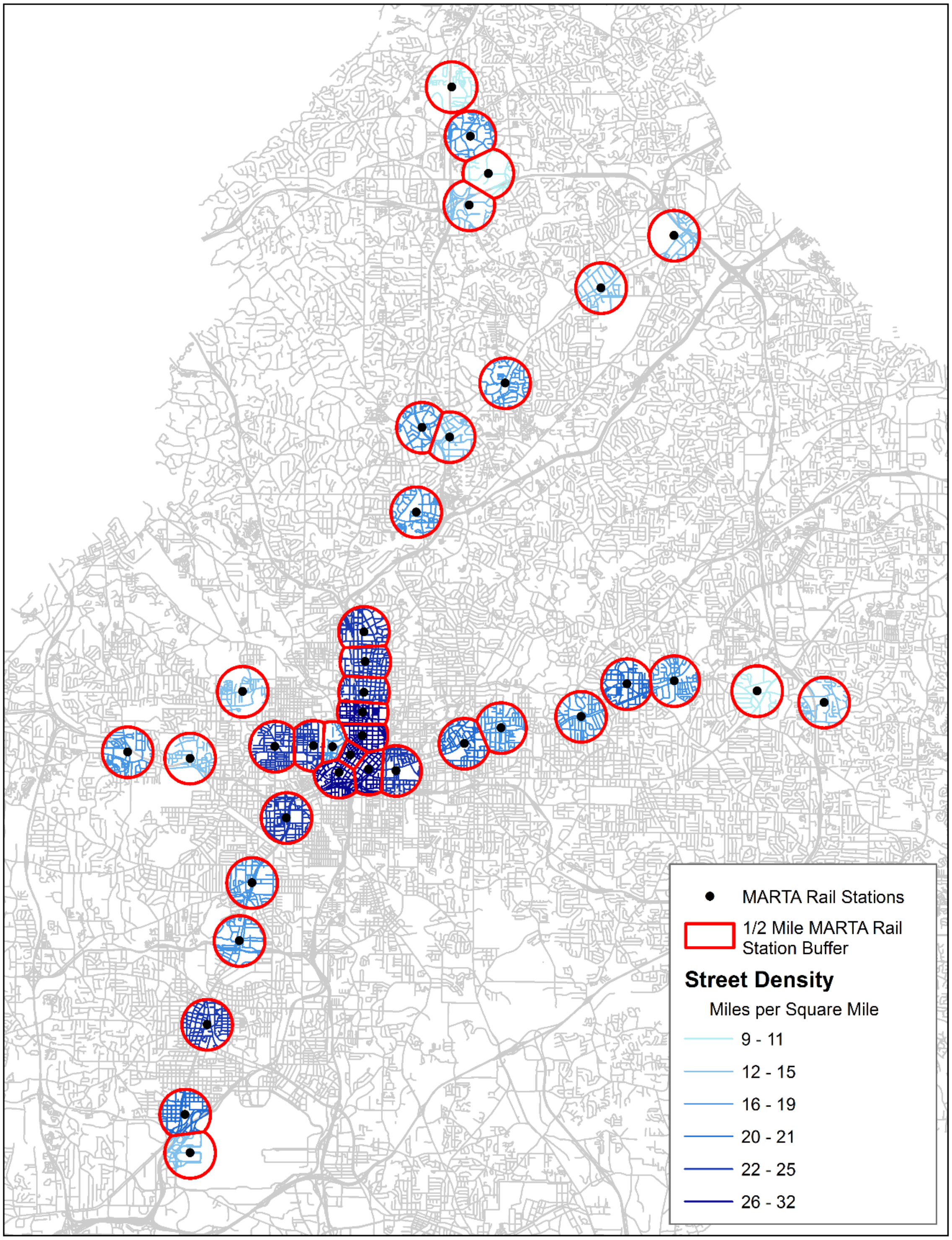

Figure 24: Street density around MARTA transit catchments

The parking variable represents the total number of parking spaces at each station. A total of fifteen out of the thirty-eight stations had no parking at all. Of the remaining eighteen stations, the greatest number of parking spaces was observed at Lindbergh with 2,907, followed by North Springs with 2,325, Kensington with 1,946, and College Park with 1,883. Besides the Airport station, the end of line stations consistently had some of the greatest number of parking spaces. Conversely, there were more stations in the Downtown area that did not have any MARTA parking (there is still some neighborhood parking available near these stations), which occurred less often on the periphery stations.

4.3 Decision tree analysis

Decision tree analysis was performed to reveal the most influential independent variables, and to identify important splits in the variables that help indicate more or less pedestrian ridership. Binary Recursive Partitioning (BRP) identifies the most important variables that influence a response variable, and subsequently splits the variables into two categories – one category leads to more of the response variable and the other category leads to less of the response variable. The identified variables in the tree each have a node that gives details of each side of the split, where the response variable increases or decreases depending on the relationship. A more detailed explanation of the utility of BRP is included in Section 3.2.

Decision tree analysis was conducted in R Studio version 0.99.465 using the rpart package to better understand the relationship of the candidate variables. Initially, all variables were loaded into R without any grouping of variables, resulting in small decision trees with variables such as the Native Hawaiian variable in the tree with extremely small values (such as splits at 0.1601). For practicality reasons, it was not desirable to include variables with extremely small counts in each of the stations. The small numbers for some categories such as "Mining, Quarrying, and Oil and Gas Extraction" in employment may help BRP come to a better predictive model for pedestrian ridership, however these categories are better left out for the generalizability of the model. Therefore grouping and eliminating of some of the smallest variables narrowed down the list of candidates to a list of 27 variables, shown below in Table 17. Additional output statistics of the decision tree analysis is shown in Table 18.

As in other modeling techniques, increasing the number of parameters can improve the models ability to predict a given dataset. In BRP, adding splits to the decision tree will improve the amount of variation explained in the response variable. Adding splits to a decision tree may decrease the generalizability of the model, and therefore reduce the ability of the model to predict future data (Merkle and Shaffer 2011).

Table 17: Variables included in BRP.

Variable Type	Number of Variables	Variables
Population	4	Population total, population density, housing, housing density
Employment	6	Total jobs, job density, retail jobs, education jobs, health care and social assistance jobs, manufacturing jobs
Annual Income	3	Less than $40,000 a year; $40,000 to $75,000; greater than $75,000
Race/Ethnicity	5	Hispanic origin, not Hispanic origin, White alone, Black, Asian alone
Gender	2	Male, female
Built environment	5	Street density, number of intersections, intersection density, number of major airports, LUMI
Station Characteristics	2	Number of parking spaces, end-of-line stations

Table 18: Output statistics from BRP in R

	CP	Number of Splits	Relative Error	Misclassification Error	Standard Error
1	0.63658367	0	1.0000000	1.036224	0.5087122
2	0.05241563	1	0.3634163	1.413082	0.6133277
3	0.01000000	2	0.3110007	1.199304	0.5195405

The resulting tree from the BRP is shown below in Figure 25 below, with the decision tree showing employment density as the most influential variable in pedestrian based ridership. The decision tree shows true statements (shown below with a “yes”) along the left stems and false statements (shown with a “no”) down the right stems. Decision trees end with nodes at the bottom of the tree, which are called terminal nodes and represent a value for the response variable. The tree indicates pedestrian ridership of about 4,682 per station for employment densities of greater than 52 jobs per acre. For stations that had less than 52 jobs per acre, the tree identifies two additional splits that further define the groups of stations that fall into different categories. The first split in the category of employment of less than 52 jobs per acre is the parking variable. The model shows that stations with fewer than 14 parking spaces have greater pedestrian accessed ridership – a literal interpretation of the tree would be that rail stations that do not have greater than or equal to 14 parking spaces have on average 2,078 pedestrian accessed ridership. The side of the tree with a greater number of parking spaces is split by another variable into two more groups: stations that have a population greater than 636 that makes less than $40,000 year and stations that have a population less than or equal to 636 that do not make less than $40,000. In other words, the split to the left indicates a station with smaller low income population that has less pedestrian based ridership, and the split to the right indicates a station with a larger low income population that has much greater pedestrian based ridership.

Overall, the decision tree analysis shows that out of the 30 candidate variables, employment density has the greatest influence on pedestrian ridership. Of the stations that have lower density employment, parking spaces was the most important variable.

Stations that have both low employment density and parking available decreases the pedestrian based ridership, while having low employment density but not as much parking available increases pedestrian based ridership at stations. Furthermore, stations that have low employment density, less parking available, and small low income population, pedestrian ridership decreases. The same group with a larger low income population increases pedestrian ridership at stations.

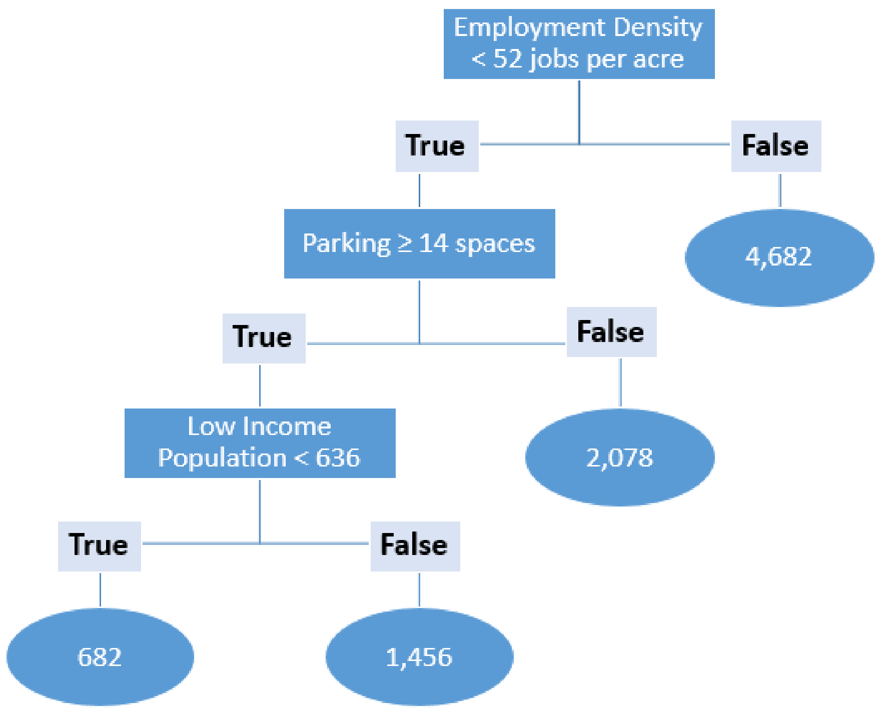

Figure 25: Decision tree with candidate variables.

4.4 DRM construction and results

The large-scale ARC on-board survey analysis in conjunction with the decision tree analysis yielded insight into the candidate variables in the previous sections. The list of ACS data variables were grouped and culled based on results and comparisons with the ARC survey, and included with the environmental and station characteristics variables. The same list of twenty-seven candidate variables used in the BRP was used in Statistical Package for the Social Sciences (SPSS) version 22, shown previously in Table 17.

Results from the comparison of transit catchment data indicated that the Airport Station had a much different set of characteristics and values than the other stations in the system. The absence of an access mode of "connected from an airport" in the ARC on-board survey likely resulted in those riders to indicate walk access rather than by another mode, evidenced by the very large number of pedestrian boardings at the Airport Station. Inclusion of a dummy variable that identified the Airport Station was considered and even tested with OLS regression, but resulted in extremely high Beta coefficients accounting for the high pedestrian based ridership at this station. For the purposes of the DRM, the Airport Station therefore was removed due to the unique environment. The Airport Station is different enough from the other stations that a completely different model is recommended to account for the drivers of ridership at this particular station.

To reduce the large number of variables for the other thirty-seven stations down to a small list to be incorporated into the model, bivariate correlations were calculated and shown in Appendix B using Pearson's correlation coefficients and level of significance (2-tailed). Many of the variables were not well correlated with the response variable, pedestrian ridership, and were dropped from consideration in the model. In addition, the correlation coefficients were utilized to remove variables that were highly correlated to reduce the likelihood of multicollinearity in the model. Each variable introduced in the model needed to be independent from all other variables. For example, the street density, the number of intersections, and the intersection density variables are all highly correlated, significant at the 0.01 level. Out of the three variables, intersection density had the highest correlation with pedestrian ridership, with a Pearson's correlation coefficient of 0.616, significant at the 0.01 level. Intersection density was therefore retained as a candidate variable for modeling purposes and street density and the number of intersections variables were removed. Additionally, variables that could be computed by the use of other variables were removed from further OLS regression to further reduce the likelihood of interdependent variables.

Of the remaining candidate variables, shown with descriptive statistics in Appendix C, employment density had by far the greatest Pearson's Correlation Coefficient with pedestrian based ridership (0.915), and was significant at the 0.01 level. Intersection density was also relatively correlated with pedestrian based ridership, which as previously mentioned, was 0.616 at the 0.01 significance level. The remaining candidate variables did not have nearly as strong of a relationship with pedestrian based ridership. The under 18 age group variable and the LUMI variable each had a negative correlation coefficient, at -0.397 and -.0353, both siginifant at the 0.05 level.

In pursuit of creating a model that can be used for future TOD development in the Atlanta area, focus was maintained on the interpretability and ease of use. Initial modelling resulted in very negative y-intercepts. While the mathematical interpretation may yield valid results, the model becomes difficult to conceptualize and interpret. It was therefore decided not to use an intercept value in the OLS regression. The generalizability of the model was also important for future use on MARTA stations and development. With only thirty-seven observations in the dataset, it was necessary to restrict the number of variables in the model. This is because a small number of observations in a dataset can inflate the coefficient of determination and reduce the generalizability of the model if a large number of predictor variables are used. Moreover, it was preferred to retain variables in the model that can be applied to policy decisions for TOD.

A stepwise regression with the final list of candidate variables with stepping method criteria of probability of F with an entry of 0.25 and a removal of 0.05 yielded three models shown in Table 19. The first model (Model 1) shown includes only the employment density variable, significant at the .05 level. The stepwise regression then added in the next most predictive variable into the model, producing Model 2, which included the low income variable (annual household income of less than $40,000). The third model added in the retail employment variable, which signifies the total number of jobs within the catchment. The employment density variable in each of the models has marginal variation, from 48.6 in the Model 1 to 44.4 in Model 2.

Table 19: DRM model coefficient summary

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	Employment Density	48.686	2.947	.940	16.518	.000
2	Employment Density	44.394	2.682	.857	16.555	.000
	Income < $40,000	.900	.222	.210	4.061	.000
3	Employment Density	47.577	2.678	.919	17.763	.000
	Income < $40,000	1.021	.206	.239	4.959	.000
	Employment (Retail)	-.336	.117	-.147	-2.879	.007

Table 20 shows the amount R-squared values for each of the three models produced in stepwise regression. The adjusted R-squared improves the amount of variation explained from 88.0% to 91.6% from Model 1 to Model 2, and then up to 93.1% in Model 3. When observing the coefficients for each of the models however, the negative sign for the employment in retail in Model 3 is unexpected. Intuitively, an increase in retail jobs would likely increase the attractiveness of a given rail station, and lead to greater ridership. The coefficient for this variable is also very small, leading to a very minor difference in ridership. Model 2 however, does not include this variable and still explains a large portion of the variation in the data.

Table 20: Stepwise model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.940	.883	.880	934.11395
2	.960	.921	.916	781.06565
3	.968	.936	.931	710.57712

Model 2 was chosen as the recommended DRM for predicting pedestrian based ridership, and is referred to as "DRM" hereafter. Descriptive statistics for the variables included in the DRM are shown in Table 21 below. The retained variables are readily available for incorporation into a simple model to predict the increase (or decrease) of ridership in modification of rail stations depending on development in the station catchment. Pearson's Correlation Coefficients are also presented in Table 22, to show how closely the variables are numerically related, and to look for multicollinearity that would artificially improve the model's prediction power. The correlations show that employment density and low income are not very correlated, with a Pearson's Correlation Coefficient of 0.394.

Table 21: DRM variable descriptive statistics

	Mean	Root Mean Square	N
Pedestrian Ridership	1724.8194	2698.81646	37
Employment Density	28.9523	52.10245	37
Income < $40,000	561.4590	630.51871	37

Table 22: Pearson's Correlation Coefficints of variables include in DRM

		Pedestrian Ridership	Employment Density	Income < $40,000
Std. Cross-product	Pedestrian Ridership	1.000	.940	.548
	Employment Density	.940	1.000	.394
	Income < $40,000	.548	.394	1.000
Sig. (1-tailed)	Pedestrian Ridership	.	.000	.000
	Employment Density	.000	.	.008
	Income < $40,000	.000	.008	.

The results from the model explained an extremely high proportion of the variance, shown in Table 23, with an adjusted R-squared of 0.916. Table 24 shows the results from an analysis of variance (ANOVA) of the DRM. The very high F-value, with statistical significance at the 0.01 level shows that the model is a good fit of the data and that the likelihood of the results happening by chance alone is extremely rare.

Table 23: DRM model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
2	.960[a]	.921	.916	781.06565

a. Predictors: Employment Density, Income < $40,000

Table 24: DRM ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
2	Regression	248141355.811	2	124070677.906	203.373	.000
	Residual	21352224.453	35	610063.556		
	Total	269493580.264	37			

Table 21 shows the coefficients for each of the variables in the model. Both of the variables were statistically significant at the 0.01 level. Both of the beta coefficients have the expected signs in the model, that is, the variables have a positive effect on pedestrian based ridership. For employment density the beta coefficient was 44.4, while income less than $40,000 had a beta coefficient of 0.9. The beta coefficients can be interpreted as the increase of pedestrian based ridership per one unit increase in predictor variable. For example, an increase of one job per acre around one of the rail stations will, on average, lead to about 44 more pedestrian boardings per day at that particular station. Increases in the number of people with an annual household income of less than $40,000 by one person per acre in the transit catchment should lead to about one more passengers accessing the rail station by walking. Most stations have transit catchments around 450 acres to 500 acres of surrounding land. The model can therefore be interpreted to yield approximately 44 more pedestrian passengers per day if the population increases about 450 or 500 jobs in the transit catchment.

Table 25: DRM model coefficient summary

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	Employment Density	44.394	2.682	.857	16.555	.000
	Income < $40,000	.900	.222	.210	4.061	.000

Figures 26 and 27 show each of the independent variables plotted against the dependent variable. Figure 26 show a clear positive linear trend for employment density and pedestrian accessed ridership, while Figure 27 depicts a similar trend, although less direct. The relationship of household income less than $40,000 annually appears to be loosely linear with the pedestrian based ridership dependent variable in the model.

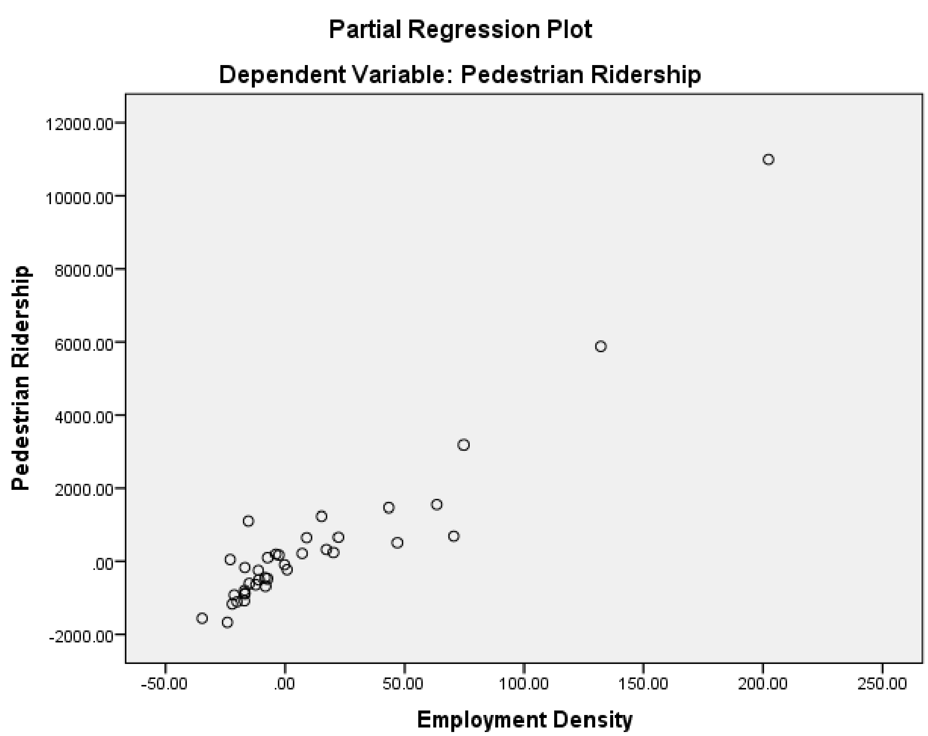

Figure 26: Plot of standardized residuals against standardized predicted values

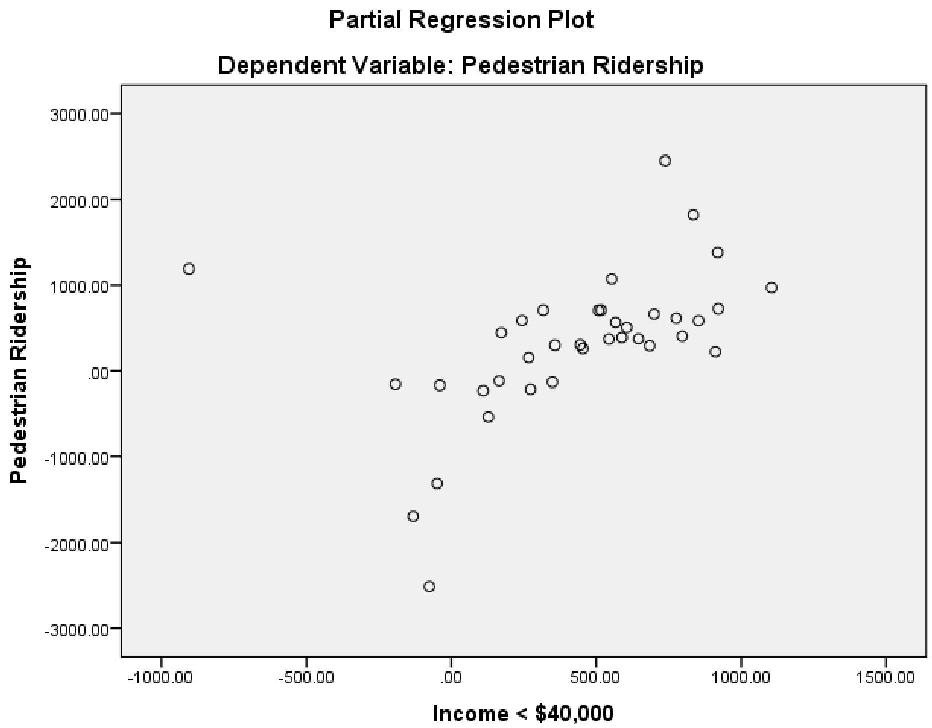

Figure 27: Plot of standardized residuals against standardized predicted values

An important step in assessing the model's generalizability is to check that the assumptions of linear regression are not violated in the model. Figure 28 shows a plot of standardized residuals against standardized predicted values to check for heteroscedasticity in the data. The plot shows that, while the number of observations limits the conclusions, the data are approximately evenly distributed above and below zero. The plot reveals that the assumptions of linearity and homoscedasticity are likely met, and that the application of the model to future observations may yield reasonable predictions of walk access ridership.

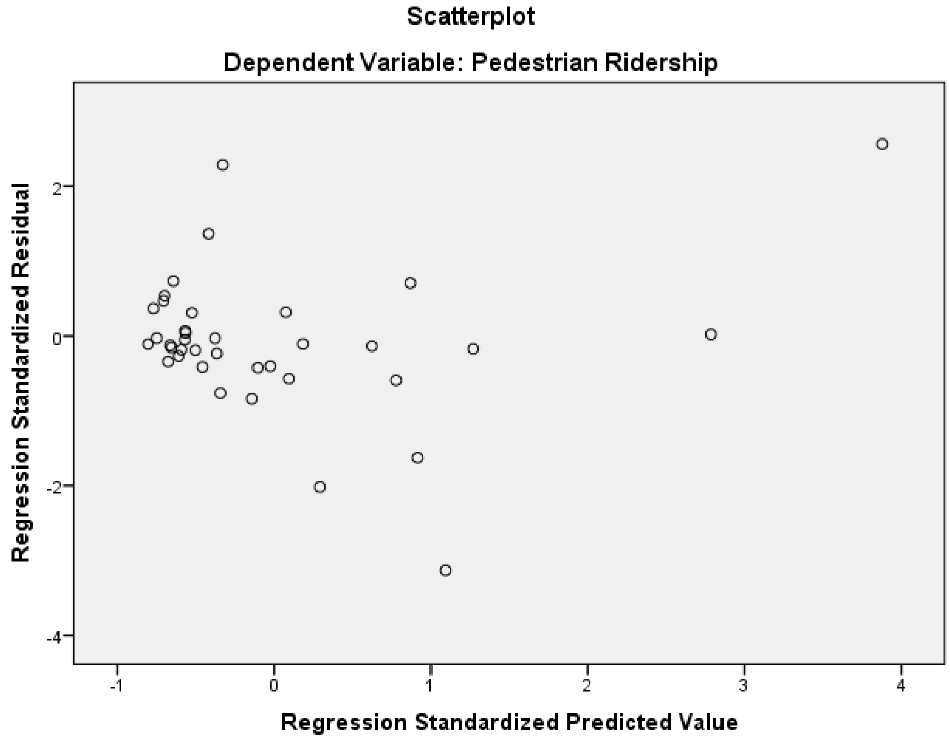

Figure 28: Plot of standardized residuals against standardized predicted values

The histogram shown below, in Figure 29, shows the frequency of the standardized residuals in the model. The bell curve shown in the plot, shows that the distribution of the residuals is relatively normal. Immediately following is Figure 30, which shows the normal probability plot, revealing a slight deviation from the line. While it is not expected to have perfectly scattered plots of standardized residuals and standardized predicted values and a perfectly alignment of normal probability plots, the plots are non-normal enough to recommend investigating if a transformation of the variables would lead to a more accurate model of walk access ridership.

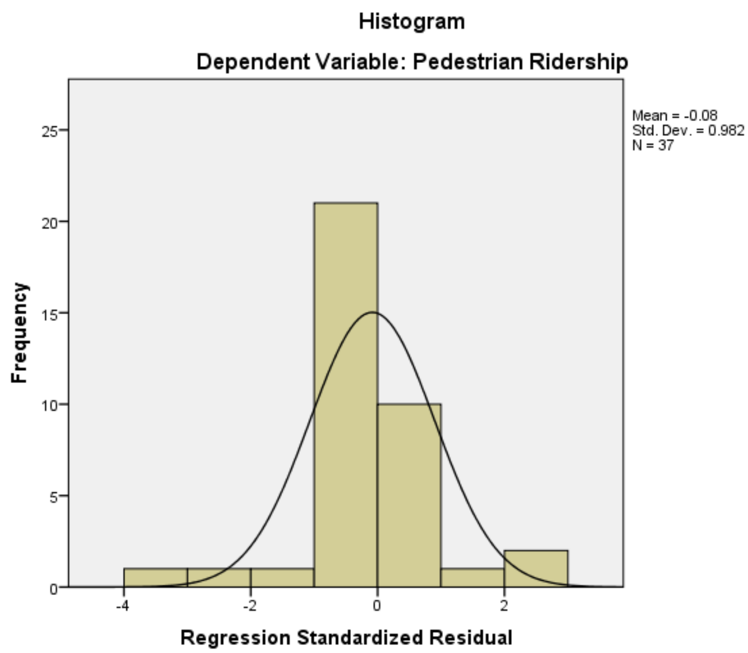

Figure 29: Histogram of distribution of residuals

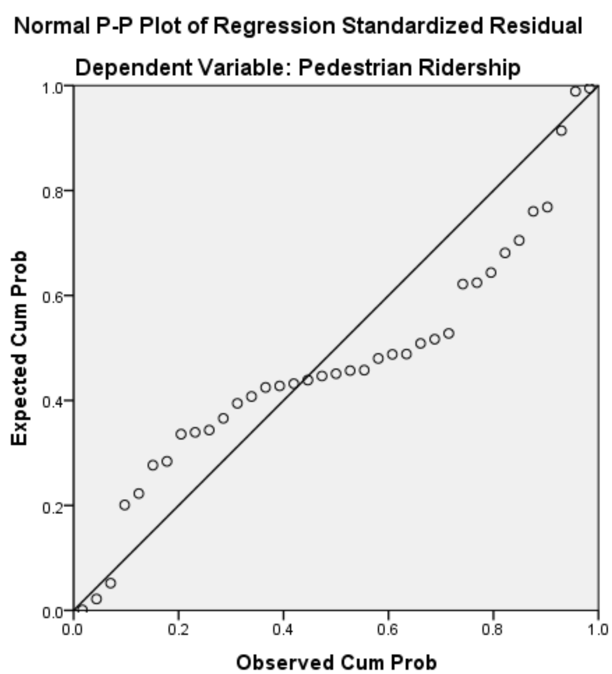

Figure 30: Normal probability plots of residuals

CHAPTER 5

CONCLUSIONS

A better understanding of pedestrian based ridership at Metropolitan Atlanta Rapid Transit Authority (MARTA) rail stations has been gained through comparison of the large scale 2010 Atlanta Regional Commission (ARC) survey and American Community Survey (ACS) data, statistical analysis of transit catchment data, and construction of a direct ridership model (DRM). The number of vehicles available showed that riders walking to the station have remarkably similar proportions of vehicle ownership to the overall population taking MARTA rail. This relationship leads one to believe that lack of vehicles is not the greatest predictor of pedestrian access to rail. Instead, the survey shows a disproportionate number of riders coming to rail by bus when they do not own vehicles. It is possible that those walking to rail have the resources to choose to live close enough to rail to make walking an easy option, and are still relatively likely to own a vehicle.

Additionally, the results from the ARC survey corroborates the background literature that lower income groups tend to take transit more often than higher income groups. This was especially true for riders accessing the rail stations via bus. Although a surprisingly high percentage of high income riders choose to walk to the stations to access rail, this appears to largely a combination of high income riders living close to MARTA rail stations and very few high income riders taking the bus to rail. MARTA transit oriented development (TOD) could benefit from providing affordable housing options close to the rail stations so that they can capitalized on a demographic that is more inclined to take transit.

The age of transit riders was also analyzed, which showed a disproportionate percentage of the population that is in the 18 to 44 age group takes rail transit. Moreover, this same group is also more likely to walk to transit than other age groups. Although the evidence in Atlanta states that these age groups are more likely to ride rail transit than others, it is important to consider to mobility options of those who are too young or too old to drive automobiles. Therefore, for the purposes of constructing a DRM that reveals ridership levels the age group of 18 to 44 could indicate increased ridership. Policy considerations however, should not focus soley on these age groups because of the equity and mobility needs of the young and elderly.

Although the gender and ethnicity variables were binary, each showed distinct characteristics that reveal preferences within categories. For gender, the results from comparisons of the ARC survey and ACS data showed that more males live close to rail stations and subsequently more males walk to rail stations to take rail transit. Conversely, the female population appears to take buses more frequently to access rail stations. Examination of the ethnicity datasets showed that those with Hispanic origin were more likely to walk to rail stations than by any other mode. Those with Hispanic origin were also more likely to walk to rail stations than those who were not of Hispanic origin. Inclusion of an ethnicity variable in a DRM could misinform the model, because the Hispanic population walking to the stations would not necessarily be captured in the transit catchment and included in the model.

Lastly, analysis on race from the ARC survey showed that the Black population was taking the bus to get to rail more often that other races. The ACS data however, showed that even though similar numbers of White and Black populations live near rail stations, the Black population was utilizing the rail stations at a much greater rate than the White population. Using insights gained from ACS data and ARC survey data together, the race variable appeared to be one of the most important variables for determining ridership numbers at rails stations, both for the all modes category as well as the pedestrian access category. Because of the extremely small values for several of the race categories, only the “White alone”, “Black” and “Asian alone” categories were included as candidate variables in the DRM.

Ultimately, the results from the comparison of the ARC survey to the ACS data revealed that using 0.5 mile buffer transit catchment ACS data in the DRM in the Atlanta area may introduce biases into the model. The ACS data is not entirely representative of the population walking to access the MARTA stations, and therefore will not necessarily lead to an accurate DRM to forecast future ridership. Specific groups such as those making less than $40,000, the age group of 18 to 44, males, Hispanic-origin ethnicity, and the African American population all show disproportionately high ridership compared to the population living in transit catchments. Special attention to these categories in the variables and inclusion into the DRMs may increase the model’s predicative power. In the case of the DRM constructed in this thesis however, these variables were not well correlated with pedestrian ridership and were not included in the model. Other, more general variables such as total population density and employment density were chosen instead to increase the generalizability and to minimize over specification. The results from the analysis of the ARC survey with ACS data are still valuable, and should be used to inform policy decisions such as mixed income housing near transit stations to offer multi-modal options to targeted populations. A strategy that includes mixed-income and focuses on low incomes (especially those below $40,000 annually) in TOD, as well as a diverse population could provide easier access to those who are more likely to utilize the transit service and could increase transit ridership.

The results from the decision tree analysis indicated that employment density is the most important factor to predict pedestrian based ridership. Employment has been shown to be the most influential variable on transit ridership in other studies, and its influence on the Atlanta transit ridership is shown here as well. By splitting the lower employment stations further by parking availability, the binary recursive partitioning (BRP) showed that having numerous parking spaces at stations is linked with lower pedestrian based ridership. This finding may suggest that having large parking lots encourages vehicular connections to rail stations, either by easy access with vehicles or by making pedestrian access difficult. Inclusion of the parking variable in the ordinary least squares (OLS) modeling however, showed a slightly positive relationship with pedestrian based ridership, which was not statistically significant.

Utilizing OLS regression as a tool to create a DRM for MARTA yielded several conclusions as well. Employment density was consistently the most influential and significant in predicting pedestrian based ridership. An increase of one job per acre in the transit catchment leads to about 44 more pedestrian based riders per day. Although the size of some transit catchments were smaller to avoid transit catchment overlap and double-counting of populations within close proximity to multiple stations, the DRM showed that most stations would need to increase the predictor variables by about 450 to 500, to show increase pedestrian based ridership by the respective coefficient. In other words, an increase of about 450 – 500 jobs in a transit catchment would likely lead to an increase of 44 riders per day. For the low income variable (households making less than $40,000 annually), the interpretation is more straightforward: an increase of about one low income person living in a transit catchment leads to about one more MARTA rider walking to the station.

The high predictive power of the DRM provides evidence that the application future TOD may be viable. In order to ascertain the efficaciousness of the model, the TOD in Atlanta needs to be developed with accurate counts of the demographic variables input into the model and tested against ridership results from pedestrian access. These values could be inserted into the model to evaluate how close the ridership numbers are to the predicted counts. Alternatively, Lindbergh Center station could serve as a case study for the DRM. Beginning in the early 2000's, commercial and residential property was developed to introduce TOD to the area. Future research could collect the demographic data of before and after the greatest developments and test these values in the model to obtain ridership estimates. The resulting ridership numbers could then be compared to pedestrian boardings data, if available, to evaluate the model's ability to predict the pedestrian ridership.

In addition to testing the DRM before and after TOD, a comparison of the results to the output from a traditional four-step travel demand model (TDM) would provide further insight into the ability of DRMs in forecasting ridership. Investigating the results from the DRM created in this thesis with the ARC TDM results would help reveal which methodology produces more accurate ridership forecasts. Analyzing ridership results at the station level would further reveal the strengths of each approach.

Some technical refinements on the modeling process is also warranted. Several of the variables investigated in the cross tabulations of riders show that the distribution of riders is not even among different demographic groups. Although several trends were revealed from the on-board survey analysis, only the demographic of household income less than $40,000 was supported in the DRM. One potential reason for some of the demographic groups failing in the DRM could be the interactions among different groups. For example, the relationship of the low income population may be more complex than a simple linear relationship with riding transit. It is possible that a certain subset of the low income population is more likely to walk to transit than other populations. Perhaps the low income population that is also in the age 18 to 44 age group walks to transit more often than the low income population that is in the 45 to 64 age group. Interactions of variables is likely the next step in refining the DRM to discover the populations that walk to transit more often.

To refine the dependent variable in the model, future work could separate the outbound and inbound trips and create two separate models. The outbound trip via walk access to the rail station would more likely be influenced by housing densities. The outbound trip would also be impacted by other modes because more riders have access to personal vehicles when departing their home. The inbound trip, would instead be heavily influenced by employment. Specifying each trip type separately would likely improve each model's performance, and yield greater insight into the relationship of pedestrians and transit.

Lastly, additional research into the transit catchment size and shape is warranted based on background literature. Although the research presented in this thesis showed a relationship with a simple buffer catchment of 0.5 miles around each rail station, the other catchment schemes could offer benefits to the creation of DRM inputs resulting in more accurate models. The potential benefits of the other more sophisticated approaches will become more attractive as the techniques involved are more automated and easily accessible by a wider audience.

The DRM presented in this thesis showed that modeling pedestrian ridership is feasible using relatively simple techniques and readily available data. Although caveats exist, planners may benefit from utilizing this technique to model alternatives of development and station location to weigh the outcomes with investments. With the increasing interest and investment in TOD in the U.S., DRMs may provide additional insight and direction into the complex options in the development landscape.

APPENDIX A MARTA RAIL STATIONS AND MODE OF ACCESS

Station	Rode in a vehicle for part of the trip and walked/biked the rest of the way	Was dropped off at a bus/train station	Carpooled/vanpooled with others and parked near the bus stop/train station	Drove alone and parked near the bus stop/train station	Walked all the way to the bus stop/train station	Bicycled all the way to the bus/train	Arrived at the station by Bus	Total
Five Points	4	235	5	53	1,363	2	1,025	2,687
GA State	1	20	1	6	481		38	547
King Memorial		35		3	68		137	243
Inman Park		33	6	53	75	3	167	337
Edgewood/Candler		27	1	42	53	1	34	158
East Lake	1	17		34	41	1	34	128
Decatur		83	7	7	163		242	502
Avondale	1	61	3	73	95	3	336	572
Kensington	9	99		163	142		649	1,062
Indian Creek	3	120	6	228	22		307	686
Peachtree Center		24	1	9	976		42	1,052
Sandy Springs	5	52	8	95	94	1	32	287
North Springs	1	70	7	282	73	1	212	646
Civic Center		18	4	3	201		31	257
North Avenue		85	9	10	473	2	112	691
Midtown		91	14	8	312	6	139	570
Art Center		101	27	8	319	2	289	746
Lindbergh	4	103	2	130	288	3	339	869
Buckhead	1	47	2	8	157		40	255
Medical Center	5	25	5	45	82		20	182
Dunwoody		43	5	58	132		112	350
Doraville	1	115	7	164	23	1	196	507
Lenox	8	35	6	45	290	2	88	474
Brookhaven	1	64	3	54	78	1	79	280
Chamblee	3	63	11	117	62	2	129	387
Bankhead	1	42		8	97		101	249

Garnett	1	9	1	1	153	1	11	177
West End	4	101	5	36	217	4	352	719
Oakland City	2	48	1	49	55	1	378	534
Lakewood/Fort McPherson		56	1	129	84	1	103	374
East Point		59	2	127	110	1	233	532
College Park	2	141	3	277	83		424	930
Airport	7	129	11	33	698		279	1,157
Dome/GWCC/Phillips/CNN		5		1	231	2	13	252
Vine City		22	1	10	69		93	195
Ashby		30	1	3	154	2	119	309
West Lake		25		10	47	1	183	266
Hamilton E Holmes		119	2	165	181	4	664	1,135
Total	65	2,452	168	2,547	8,242	48	7,782	21,304

APPENDIX B PEARSON CORRELATION OF VARIABLES

		Pedestrian Ridership	Population Density	Housing Density	Employment (Manufacturing)	Employment (Retail)
Pedestrian Ridership	Pearson Correlation	1	.340*	.361*	0.216	0.096
	Sig. (2-tailed)		0.039	0.028	0.2	0.572
Population Density	Pearson Correlation	.340*	1	.894**	-0.015	-0.1
	Sig. (2-tailed)	0.039		0	0.929	0.555
Housing Density	Pearson Correlation	.361*	.894**	1	-0.078	0.052
	Sig. (2-tailed)	0.028	0		0.648	0.76
Employment (Manufacturing)	Pearson Correlation	0.216	-0.015	-0.078	1	0.173
	Sig. (2-tailed)	0.2	0.929	0.648		0.307
Employment (Retail)	Pearson Correlation	0.096	-0.1	0.052	0.173	1
	Sig. (2-tailed)	0.572	0.555	0.76	0.307	
Employment (Education Services)	Pearson Correlation	.351*	.506**	.333*	-0.128	-0.019
	Sig. (2-tailed)	0.033	0.001	0.044	0.449	0.912
Employment (Health Care and Social Asssistance)	Pearson Correlation	0.09	-0.032	-0.071	0.066	0.103
	Sig. (2-tailed)	0.597	0.853	0.675	0.698	0.543
Employment Density	Pearson Correlation	.915**	0.187	0.227	0.295	0.292
	Sig. (2-tailed)	0	0.267	0.178	0.076	0.08
Income < $40,000	Pearson Correlation	-0.06	.729**	.624**	-0.099	-0.263
	Sig. (2-tailed)	0.723	0	0	0.56	0.115
Income $40,000 to $75,000	Pearson Correlation	0.061	.761**	.824**	-0.027	0.087
	Sig. (2-tailed)	0.72	0	0	0.876	0.61
Income > $75,000	Pearson Correlation	0.052	.562**	.732**	0.031	0.309
	Sig. (2-tailed)	0.761	0	0	0.854	0.063
Hispanic Ethnicity	Pearson Correlation	0.043	.895**	.804**	-0.099	-0.144
	Sig. (2-tailed)	0.801	0	0	0.56	0.395
Non-hispanic Ethnity	Pearson Correlation	0.015	0.23	0.285	0.003	0.041
	Sig. (2-tailed)	0.93	0.171	0.087	0.984	0.81
White Alone	Pearson Correlation	0.109	.701**	.801**	0.006	0.157
	Sig. (2-tailed)	0.522	0	0	0.973	0.354
Black	Pearson Correlation	-0.095	.341*	0.102	-0.17	-.401*
	Sig. (2-tailed)	0.574	0.039	0.55	0.315	0.014
Asian Alone	Pearson Correlation	0.127	.615**	.615**	0.086	0.093
	Sig. (2-tailed)	0.454	0	0	0.611	0.584
Age < 18	Pearson Correlation	-.397*	0.116	0.007	-0.258	-.336*
	Sig. (2-tailed)	0.015	0.495	0.966	0.123	0.042
Age 18 to 44	Pearson Correlation	0.193	.941**	.864**	-0.011	-0.061
	Sig. (2-tailed)	0.253	0	0	0.946	0.718
Age 45 to 64	Pearson Correlation	-0.177	.682**	.661**	-0.195	-0.207
	Sig. (2-tailed)	0.294	0	0	0.249	0.218
Age 65 and up	Pearson Correlation	-0.189	0.246	0.261	-0.145	0.013
	Sig. (2-tailed)	0.262	0.142	0.119	0.393	0.94
Males	Pearson Correlation	0.069	.918**	.844**	-0.119	-0.129
	Sig. (2-tailed)	0.683	0	0	0.482	0.448
Females	Pearson Correlation	0.001	.756**	.681**	-0.047	-0.121
	Sig. (2-tailed)	0.996	0	0	0.782	0.476
Street Density	Pearson Correlation	.560**	.483**	.433**	0.066	-0.214
	Sig. (2-tailed)	0	0.002	0.007	0.7	0.204
Intersections	Pearson Correlation	0.086	.375*	0.286	0.07	-0.302
	Sig. (2-tailed)	0.612	0.022	0.086	0.678	0.069
Intersection Density	Pearson Correlation	.616**	.479**	.397*	0.101	-0.22
	Sig. (2-tailed)	0	0.003	0.015	0.551	0.191
LUMI	Pearson Correlation	-.353*	-0.122	-0.076	-0.122	-0.313
	Sig. (2-tailed)	0.032	0.471	0.655	0.472	0.059
Parking	Pearson Correlation	-0.298	-0.276	-0.254	-0.158	-0.116
	Sig. (2-tailed)	0.074	0.098	0.129	0.352	0.495
End of Line	Pearson Correlation	-0.214	-0.197	-0.22	0.046	-0.125
	Sig. (2-tailed)	0.203	0.243	0.19	0.788	0.461

		Employment (Education Services)	Employment (Health Care and Social Asssistance)	Employment Density	Income < $40,000	Income $40,000 to $75,000
Pedestrian Ridership	Pearson Correlation	.351*	0.09	.915**	-0.06	0.061
	Sig. (2-tailed)	0.033	0.597	0	0.723	0.72
Population Density	Pearson Correlation	.506**	-0.032	0.187	.729**	.761**
	Sig. (2-tailed)	0.001	0.853	0.267	0	0
Housing Density	Pearson Correlation	.333*	-0.071	0.227	.624**	.824**
	Sig. (2-tailed)	0.044	0.675	0.178	0	0
Employment (Manufacturing)	Pearson Correlation	-0.128	0.066	0.295	-0.099	-0.027
	Sig. (2-tailed)	0.449	0.698	0.076	0.56	0.876
Employment (Retail)	Pearson Correlation	-0.019	0.103	0.292	-0.263	0.087
	Sig. (2-tailed)	0.912	0.543	0.08	0.115	0.61
Employment (Education Services)	Pearson Correlation	1	0.29	0.292	0.268	0.32
	Sig. (2-tailed)		0.082	0.079	0.109	0.054
Employment (Health Care and Social Asssistance)	Pearson Correlation	0.29	1	0.237	-0.192	-0.053
	Sig. (2-tailed)	0.082		0.158	0.256	0.754
Employment Density	Pearson Correlation	0.292	0.237	1	-0.266	-0.077
	Sig. (2-tailed)	0.079	0.158		0.111	0.651
Income < $40,000	Pearson Correlation	0.268	-0.192	-0.266	1	.606**
	Sig. (2-tailed)	0.109	0.256	0.111		0
Income $40,000 to $75,000	Pearson Correlation	0.32	-0.053	-0.077	.606**	1
	Sig. (2-tailed)	0.054	0.754	0.651	0	
Income > $75,000	Pearson Correlation	0.137	0.009	0.023	0.322	.865**
	Sig. (2-tailed)	0.418	0.96	0.893	0.052	0
Hispanic Ethnicity	Pearson Correlation	.333*	-0.126	-0.121	.838**	.798**
	Sig. (2-tailed)	0.044	0.457	0.476	0	0
Non-hispanic Ethnity	Pearson Correlation	0.01	-0.121	-0.137	0.099	.378*
	Sig. (2-tailed)	0.954	0.477	0.418	0.558	0.021
White Alone	Pearson Correlation	0.28	-0.031	0.038	.429**	.902**
	Sig. (2-tailed)	0.094	0.856	0.821	0.008	0
Black	Pearson Correlation	0.042	-0.161	-0.245	.643**	-0.049
	Sig. (2-tailed)	0.805	0.342	0.145	0	0.775
Asian Alone	Pearson Correlation	.455**	0.029	0.067	0.295	.740**
	Sig. (2-tailed)	0.005	0.864	0.695	0.076	0
Age < 18	Pearson Correlation	-0.22	-0.312	-.575**	.386*	0.156
	Sig. (2-tailed)	0.19	0.06	0	0.018	0.357
Age 18 to 44	Pearson Correlation	.445**	-0.054	0.032	.755**	.838**
	Sig. (2-tailed)	0.006	0.752	0.849	0	0
Age 45 to 64	Pearson Correlation	-0.013	-0.256	-0.31	.728**	.628**
	Sig. (2-tailed)	0.938	0.126	0.062	0	0
Age 65 and up	Pearson Correlation	0.092	-0.081	-0.261	.492**	.392*
	Sig. (2-tailed)	0.587	0.633	0.118	0.002	0.016
Males	Pearson Correlation	.352*	-0.11	-0.1	.802**	.802**
	Sig. (2-tailed)	0.033	0.515	0.557	0	0
Females	Pearson Correlation	0.24	-0.166	-0.176	.751**	.761**
	Sig. (2-tailed)	0.153	0.325	0.296	0	0
Street Density	Pearson Correlation	.338*	-0.009	.487**	0.281	0.092
	Sig. (2-tailed)	0.041	0.96	0.002	0.092	0.59
Intersections	Pearson Correlation	0.2	-0.151	-0.039	.536**	0.122
	Sig. (2-tailed)	0.235	0.372	0.817	0.001	0.474
Intersection Density	Pearson Correlation	.377*	-0.027	.543**	0.279	0.025
	Sig. (2-tailed)	0.022	0.874	0.001	0.094	0.883
LUMI	Pearson Correlation	-0.199	-.352*	-.487**	0.239	-0.097
	Sig. (2-tailed)	0.238	0.033	0.002	0.154	0.568
Parking	Pearson Correlation	-0.261	-0.251	-.424**	-0.212	0.034
	Sig. (2-tailed)	0.119	0.135	0.009	0.207	0.843
End of Line	Pearson Correlation	-0.154	-0.127	-0.228	-0.157	-0.127
	Sig. (2-tailed)	0.364	0.454	0.175	0.352	0.452

		Income > $75,000	Hispanic Ethnicity	Non-hispanic Ethnity	White Alone	Black
Pedestrian Ridership	Pearson Correlation	0.052	0.043	0.015	0.109	-0.095
	Sig. (2-tailed)	0.761	0.801	0.93	0.522	0.574
Population Density	Pearson Correlation	.562**	.895**	0.23	.701**	.341*
	Sig. (2-tailed)	0	0	0.171	0	0.039
Housing Density	Pearson Correlation	.732**	.804**	0.285	.801**	0.102
	Sig. (2-tailed)	0	0	0.087	0	0.55
Employment (Manufacturing)	Pearson Correlation	0.031	-0.099	0.003	0.006	-0.17
	Sig. (2-tailed)	0.854	0.56	0.984	0.973	0.315
Employment (Retail)	Pearson Correlation	0.309	-0.144	0.041	0.157	-.401*
	Sig. (2-tailed)	0.063	0.395	0.81	0.354	0.014
Employment (Education Services)	Pearson Correlation	0.137	.333*	0.01	0.28	0.042
	Sig. (2-tailed)	0.418	0.044	0.954	0.094	0.805
Employment (Health Care and Social Asssistance)	Pearson Correlation	0.009	-0.126	-0.121	-0.031	-0.161
	Sig. (2-tailed)	0.96	0.457	0.477	0.856	0.342
Employment Density	Pearson Correlation	0.023	-0.121	-0.137	0.038	-0.245
	Sig. (2-tailed)	0.893	0.476	0.418	0.821	0.145
Income < $40,000	Pearson Correlation	0.322	.838**	0.099	.429**	.643**
	Sig. (2-tailed)	0.052	0	0.558	0.008	0
Income $40,000 to $75,000	Pearson Correlation	.865**	.798**	.378*	.902**	-0.049
	Sig. (2-tailed)	0	0	0.021	0	0.775
Income > $75,000	Pearson Correlation	1	.632**	0.299	.944**	-0.313
	Sig. (2-tailed)		0	0.072	0	0.059
Hispanic Ethnicity	Pearson Correlation	.632**	1	0.187	.732**	.469**
	Sig. (2-tailed)	0		0.267	0	0.003
Non-hispanic Ethnity	Pearson Correlation	0.299	0.187	1	.358*	-0.163
	Sig. (2-tailed)	0.072	0.267		0.029	0.335
White Alone	Pearson Correlation	.944**	.732**	.358*	1	-0.247
	Sig. (2-tailed)	0	0	0.029		0.141
Black	Pearson Correlation	-0.313	.469**	-0.163	-0.247	1
	Sig. (2-tailed)	0.059	0.003	0.335	0.141	
Asian Alone	Pearson Correlation	.625**	.547**	.376*	.685**	-0.217
	Sig. (2-tailed)	0	0	0.022	0	0.196
Age < 18	Pearson Correlation	0.052	.446**	0.132	0.096	.549**
	Sig. (2-tailed)	0.758	0.006	0.436	0.573	0
Age 18 to 44	Pearson Correlation	.660**	.933**	.382*	.769**	0.316
	Sig. (2-tailed)	0	0	0.02	0	0.057
Age 45 to 64	Pearson Correlation	.552**	.858**	0.095	.617**	.450**
	Sig. (2-tailed)	0	0	0.574	0	0.005
Age 65 and up	Pearson Correlation	.354*	.479**	0.018	.383*	0.21
	Sig. (2-tailed)	0.031	0.003	0.917	0.019	0.212
Males	Pearson Correlation	.612**	.954**	.342*	.722**	.416*
	Sig. (2-tailed)	0	0	0.038	0	0.011
Females	Pearson Correlation	.637**	.925**	0.286	.724**	.389*
	Sig. (2-tailed)	0	0	0.086	0	0.017
Street Density	Pearson Correlation	0.016	0.247	-0.161	0.136	0.185
	Sig. (2-tailed)	0.926	0.14	0.34	0.423	0.272
Intersections	Pearson Correlation	-0.011	.348*	-0.123	0.097	.391*
	Sig. (2-tailed)	0.948	0.035	0.468	0.567	0.017
Intersection Density	Pearson Correlation	-0.087	0.2	-0.17	0.04	0.223
	Sig. (2-tailed)	0.61	0.235	0.313	0.813	0.185
LUMI	Pearson Correlation	-0.258	-0.04	0.272	-0.184	0.208
	Sig. (2-tailed)	0.122	0.812	0.103	0.277	0.216
Parking	Pearson Correlation	-0.062	-0.14	.457**	-0.091	-0.094
	Sig. (2-tailed)	0.716	0.407	0.004	0.591	0.578
End of Line	Pearson Correlation	-0.225	-0.128	-0.005	-0.232	0.068
	Sig. (2-tailed)	0.181	0.451	0.976	0.167	0.69

		Asian Alone	Age < 18	Age 18 to 44	Age 45 to 64	Age 65 and up
Pedestrian Ridership	Pearson Correlation	0.127	-.397*	0.193	-0.177	-0.189
	Sig. (2-tailed)	0.454	0.015	0.253	0.294	0.262
Population Density	Pearson Correlation	.615**	0.116	.941**	.682**	0.246
	Sig. (2-tailed)	0	0.495	0	0	0.142
Housing Density	Pearson Correlation	.615**	0.007	.864**	.661**	0.261
	Sig. (2-tailed)	0	0.966	0	0	0.119
Employment (Manufacturing)	Pearson Correlation	0.086	-0.258	-0.011	-0.195	-0.145
	Sig. (2-tailed)	0.611	0.123	0.946	0.249	0.393
Employment (Retail)	Pearson Correlation	0.093	-.336*	-0.061	-0.207	0.013
	Sig. (2-tailed)	0.584	0.042	0.718	0.218	0.94
Employment (Education Services)	Pearson Correlation	.455**	-0.22	.445**	-0.013	0.092
	Sig. (2-tailed)	0.005	0.19	0.006	0.938	0.587
Employment (Health Care and Social Asssistance)	Pearson Correlation	0.029	-0.312	-0.054	-0.256	-0.081
	Sig. (2-tailed)	0.864	0.06	0.752	0.126	0.633
Employment Density	Pearson Correlation	0.067	-.575**	0.032	-0.31	-0.261
	Sig. (2-tailed)	0.695	0	0.849	0.062	0.118
Income < $40,000	Pearson Correlation	0.295	.386*	.755**	.728**	.492**
	Sig. (2-tailed)	0.076	0.018	0	0	0.002
Income $40,000 to $75,000	Pearson Correlation	.740**	0.156	.838**	.628**	.392*
	Sig. (2-tailed)	0	0.357	0	0	0.016
Income > $75,000	Pearson Correlation	.625**	0.052	.660**	.552**	.354*
	Sig. (2-tailed)	0	0.758	0	0	0.031
Hispanic Ethnicity	Pearson Correlation	.547**	.446**	.933**	.858**	.479**
	Sig. (2-tailed)	0	0.006	0	0	0.003
Non-hispanic Ethnity	Pearson Correlation	.376*	0.132	.382*	0.095	0.018
	Sig. (2-tailed)	0.022	0.436	0.02	0.574	0.917
White Alone	Pearson Correlation	.685**	0.096	.769**	.617**	.383*
	Sig. (2-tailed)	0	0.573	0	0	0.019
Black	Pearson Correlation	-0.217	.549**	0.316	.450**	0.21
	Sig. (2-tailed)	0.196	0	0.057	0.005	0.212
Asian Alone	Pearson Correlation	1	-0.059	.677**	0.303	0.117
	Sig. (2-tailed)		0.728	0	0.068	0.491
Age < 18	Pearson Correlation	-0.059	1	0.206	.557**	.468**
	Sig. (2-tailed)	0.728		0.221	0	0.003
Age 18 to 44	Pearson Correlation	.677**	0.206	1	.682**	0.265
	Sig. (2-tailed)	0	0.221		0	0.113
Age 45 to 64	Pearson Correlation	0.303	.557**	.682**	1	.523**
	Sig. (2-tailed)	0.068	0	0		0.001
Age 65 and up	Pearson Correlation	0.117	.468**	0.265	.523**	1
	Sig. (2-tailed)	0.491	0.003	0.113	0.001	
Males	Pearson Correlation	.601**	0.312	.968**	.800**	0.291
	Sig. (2-tailed)	0	0.06	0	0	0.081
Females	Pearson Correlation	.491**	.594**	.824**	.797**	.658**
	Sig. (2-tailed)	0.002	0	0	0	0
Street Density	Pearson Correlation	-0.009	-0.261	0.279	0.218	-0.065
	Sig. (2-tailed)	0.959	0.119	0.094	0.194	0.704
Intersections	Pearson Correlation	-0.063	0.027	0.293	.377*	0.161
	Sig. (2-tailed)	0.712	0.872	0.078	0.022	0.34
Intersection Density	Pearson Correlation	0.015	-.346*	0.264	0.129	-0.139
	Sig. (2-tailed)	0.931	0.036	0.114	0.445	0.411
LUMI	Pearson Correlation	-0.152	0.217	-0.084	0.175	0.072
	Sig. (2-tailed)	0.369	0.198	0.62	0.3	0.673
Parking	Pearson Correlation	0.148	.343*	-0.094	-0.136	-0.166
	Sig. (2-tailed)	0.383	0.037	0.582	0.421	0.325
End of Line	Pearson Correlation	0.099	0.121	-0.152	-0.089	-0.067
	Sig. (2-tailed)	0.56	0.475	0.368	0.599	0.694

		Males	Females	Street Density	Intersections	Intersection Density
Pedestrian Ridership	Pearson Correlation	0.069	0.001	.560**	0.086	.616**
	Sig. (2-tailed)	0.683	0.996	0	0.612	0
Population Density	Pearson Correlation	.918**	.756**	.483**	.375*	.479**
	Sig. (2-tailed)	0	0	0.002	0.022	0.003
Housing Density	Pearson Correlation	.844**	.681**	.433**	0.286	.397*
	Sig. (2-tailed)	0	0	0.007	0.086	0.015
Employment (Manufacturing)	Pearson Correlation	-0.119	-0.047	0.066	0.07	0.101
	Sig. (2-tailed)	0.482	0.782	0.7	0.678	0.551
Employment (Retail)	Pearson Correlation	-0.129	-0.121	-0.214	-0.302	-0.22
	Sig. (2-tailed)	0.448	0.476	0.204	0.069	0.191
Employment (Education Services)	Pearson Correlation	.352*	0.24	.338*	0.2	.377*
	Sig. (2-tailed)	0.033	0.153	0.041	0.235	0.022
Employment (Health Care and Social Asssistance)	Pearson Correlation	-0.11	-0.166	-0.009	-0.151	-0.027
	Sig. (2-tailed)	0.515	0.325	0.96	0.372	0.874
Employment Density	Pearson Correlation	-0.1	-0.176	.487**	-0.039	.543**
	Sig. (2-tailed)	0.557	0.296	0.002	0.817	0.001
Income < $40,000	Pearson Correlation	.802**	.751**	0.281	.536**	0.279
	Sig. (2-tailed)	0	0	0.092	0.001	0.094
Income $40,000 to $75,000	Pearson Correlation	.802**	.761**	0.092	0.122	0.025
	Sig. (2-tailed)	0	0	0.59	0.474	0.883
Income > $75,000	Pearson Correlation	.612**	.637**	0.016	-0.011	-0.087
	Sig. (2-tailed)	0	0	0.926	0.948	0.61
Hispanic Ethnicity	Pearson Correlation	.954**	.925**	0.247	.348*	0.2
	Sig. (2-tailed)	0	0	0.14	0.035	0.235
Non-hispanic Ethnity	Pearson Correlation	.342*	0.286	-0.161	-0.123	-0.17
	Sig. (2-tailed)	0.038	0.086	0.34	0.468	0.313
White Alone	Pearson Correlation	.722**	.724**	0.136	0.097	0.04
	Sig. (2-tailed)	0	0	0.423	0.567	0.813
Black	Pearson Correlation	.416*	.389*	0.185	.391*	0.223
	Sig. (2-tailed)	0.011	0.017	0.272	0.017	0.185
Asian Alone	Pearson Correlation	.601**	.491**	-0.009	-0.063	0.015
	Sig. (2-tailed)	0	0.002	0.959	0.712	0.931
Age < 18	Pearson Correlation	0.312	.594**	-0.261	0.027	-.346*
	Sig. (2-tailed)	0.06	0	0.119	0.872	0.036
Age 18 to 44	Pearson Correlation	.968**	.824**	0.279	0.293	0.264
	Sig. (2-tailed)	0	0	0.094	0.078	0.114
Age 45 to 64	Pearson Correlation	.800**	.797**	0.218	.377*	0.129
	Sig. (2-tailed)	0	0	0.194	0.022	0.445
Age 65 and up	Pearson Correlation	0.291	.658**	-0.065	0.161	-0.139
	Sig. (2-tailed)	0.081	0	0.704	0.34	0.411
Males	Pearson Correlation	1	.811**	0.242	0.306	0.224
	Sig. (2-tailed)		0	0.148	0.066	0.182
Females	Pearson Correlation	.811**	1	0.144	0.291	0.061
	Sig. (2-tailed)	0		0.395	0.081	0.722
Street Density	Pearson Correlation	0.242	0.144	1	.761**	.947**
	Sig. (2-tailed)	0.148	0.395		0	0
Intersections	Pearson Correlation	0.306	0.291	.761**	1	.729**
	Sig. (2-tailed)	0.066	0.081	0		0
Intersection Density	Pearson Correlation	0.224	0.061	.947**	.729**	1
	Sig. (2-tailed)	0.182	0.722	0	0	
LUMI	Pearson Correlation	0.038	-0.047	-0.08	0.252	-0.055
	Sig. (2-tailed)	0.824	0.782	0.638	0.133	0.747
Parking	Pearson Correlation	-0.091	-0.016	-.553**	-.420**	-.575**
	Sig. (2-tailed)	0.591	0.925	0	0.01	0
End of Line	Pearson Correlation	-0.112	-0.126	-.423**	-0.226	-0.292
	Sig. (2-tailed)	0.508	0.459	0.009	0.179	0.079

		LUMI	Parking	End of Line
Pedestrian Ridership	Pearson Correlation	-.353*	-0.298	-0.214
	Sig. (2-tailed)	0.032	0.074	0.203
Population Density	Pearson Correlation	-0.122	-0.276	-0.197
	Sig. (2-tailed)	0.471	0.098	0.243
Housing Density	Pearson Correlation	-0.076	-0.254	-0.22
	Sig. (2-tailed)	0.655	0.129	0.19
Employment (Manufacturing)	Pearson Correlation	-0.122	-0.158	0.046
	Sig. (2-tailed)	0.472	0.352	0.788
Employment (Retail)	Pearson Correlation	-0.313	-0.116	-0.125
	Sig. (2-tailed)	0.059	0.495	0.461
Employment (Education Services)	Pearson Correlation	-0.199	-0.261	-0.154
	Sig. (2-tailed)	0.238	0.119	0.364
Employment (Health Care and Social Asssistance)	Pearson Correlation	-.352*	-0.251	-0.127
	Sig. (2-tailed)	0.033	0.135	0.454
Employment Density	Pearson Correlation	-.487**	-.424**	-0.228
	Sig. (2-tailed)	0.002	0.009	0.175
Income < $40,000	Pearson Correlation	0.239	-0.212	-0.157
	Sig. (2-tailed)	0.154	0.207	0.352
Income $40,000 to $75,000	Pearson Correlation	-0.097	0.034	-0.127
	Sig. (2-tailed)	0.568	0.843	0.452
Income > $75,000	Pearson Correlation	-0.258	-0.062	-0.225
	Sig. (2-tailed)	0.122	0.716	0.181
Hispanic Ethnicity	Pearson Correlation	-0.04	-0.14	-0.128
	Sig. (2-tailed)	0.812	0.407	0.451
Non-hispanic Ethnity	Pearson Correlation	0.272	.457**	-0.005
	Sig. (2-tailed)	0.103	0.004	0.976
White Alone	Pearson Correlation	-0.184	-0.091	-0.232
	Sig. (2-tailed)	0.277	0.591	0.167
Black	Pearson Correlation	0.208	-0.094	0.068
	Sig. (2-tailed)	0.216	0.578	0.69
Asian Alone	Pearson Correlation	-0.152	0.148	0.099
	Sig. (2-tailed)	0.369	0.383	0.56
Age < 18	Pearson Correlation	0.217	.343*	0.121
	Sig. (2-tailed)	0.198	0.037	0.475
Age 18 to 44	Pearson Correlation	-0.084	-0.094	-0.152
	Sig. (2-tailed)	0.62	0.582	0.368
Age 45 to 64	Pearson Correlation	0.175	-0.136	-0.089
	Sig. (2-tailed)	0.3	0.421	0.599
Age 65 and up	Pearson Correlation	0.072	-0.166	-0.067
	Sig. (2-tailed)	0.673	0.325	0.694
Males	Pearson Correlation	0.038	-0.091	-0.112
	Sig. (2-tailed)	0.824	0.591	0.508
Females	Pearson Correlation	-0.047	-0.016	-0.126
	Sig. (2-tailed)	0.782	0.925	0.459
Street Density	Pearson Correlation	-0.08	-.553**	-.423**
	Sig. (2-tailed)	0.638	0	0.009
Intersections	Pearson Correlation	0.252	-.420**	-0.226
	Sig. (2-tailed)	0.133	0.01	0.179
Intersection Density	Pearson Correlation	-0.055	-.575**	-0.292
	Sig. (2-tailed)	0.747	0	0.079
LUMI	Pearson Correlation	1	0.157	0.266
	Sig. (2-tailed)		0.353	0.112
Parking	Pearson Correlation	0.157	1	.368*
	Sig. (2-tailed)	0.353		0.025
End of Line	Pearson Correlation	0.266	.368*	1
	Sig. (2-tailed)	0.112	0.025	

* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

APPENDIX C DESCRIPTIVE STATISTICS OF DRM VARIABLES

Variable	Mean	Root Mean Square	N
Pedestrian Ridership	1724.8194	2698.81646	37
Population Density	7.4169	8.99326	37
Employment (Manufacturing)	131.5543	198.85014	37
Employment (Retail)	610.2948	1177.77437	37
Employment (Education Services)	356.2057	867.62048	37
Employment (Health Care and Social Asssistance)	927.4742	2761.94475	37
Employment Density	28.9523	52.10245	37
Income < $40,000	561.4590	630.51871	37
Income > $75,000	423.7486	572.79194	37
Hispanic Ethnicity	2789.8340	3243.58050	37
Black	1336.8957	1759.83486	37
Asian Alone	187.6707	325.40941	37
Age < 18	403.2066	473.32507	37
Age 18 to 44	1754.1949	2208.88768	37
Age 65 and up	214.9762	271.00605	37
Males	1594.6735	1921.84928	37
Intersection Density	390.4409	447.95973	37
LUMI	.5733	.61761	37
Parking	686.9189	1057.42193	37
End of Line	.1351	.36761	37

REFERENCES

Biba, S., Curtin, K. M., & Manca, G. (2010). A new method for determining the population with walking access to transit. *International Journal of Geographical Information Science*, *24*(3), 347-364.Cervero, R. (2006). Alternative approaches to modeling the travel-demand impacts of smart growth. *Journal of the American Planning Association*, *72*(3), 285-295.

Brosnan, R. (2010). 40 Years of Transit Oriented Development: Arlington County's Experience with Transit Oriented Development in the Rosslyn-Ballston Metro Corridor [PDF]. Retrieved from http://www.fairfaxcounty.gov/dpz/projects/reston/presentations/40_years_of_transit_oriented_development.pdf

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment*, *7*(4), 265-284.

Cervero, R. (2006). Alternative Approaches to Modeling the Travel-Demand Impacts of Smart Growth. *Journal of the American Planning Association*, 72(3), 285-295.

Cervero, R., Murakami, J., & Miller, M. (2010). Direct ridership model of bus rapid transit in Los Angeles County, California. *Transportation Research Record: Journal of the Transportation Research Board*, (2145), 1-7.

Chakraborty, J., & Armstrong, M. P. (1997). Exploring the use of buffer analysis for the identification of impacted areas in environmental equity assessment. *Cartography and Geographic Information Systems*, *24*(3), 145-157.

Chow, L. F., Zhao, F., Liu, X., Li, M. T., & Ubaka, I. (2006). Transit ridership model based on geographically weighted regression. *Transportation Research Record: Journal of the Transportation Research Board*, (1972), 105-114.

Chu, X. (2004). *Ridership models at the stop level* (No. NCTR-473-04, BC137-31).

Dill, J. (2004, January). Measuring network connectivity for bicycling and walking. In *83rd Annual Meeting of the Transportation Research Board, Washington, DC* (pp. 11-15).

ETC Institute, (2010). Regional On-Board Transit Survey Final Report. Olathe, Kansas: ETC Institute in association with PBS&J and DW & Associates.

Filion, P. (2001). Suburban mixed-use centres and urban dispersion: what difference do they make?. *Environment and Planning a*, *33*(1), 141-160.

Goll, D. (2015, April 27). BART Creating Communities Around Transit Stations. *The Registry Bay Real Estate*, Retrieved from http://news.theregistrysf.com

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-Mile Circle: Does It Best Represent Transit Station Catchments?. *Transportation Research Record: Journal of the Transportation Research Board*, (2276), 101-109.

Guerra, E., & Cervero, R. (2013). Is a Half-Mile Circle the Right Standard for TODs?. *ACCESS Magazine*.

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2008). Distance measure impacts of public transport service areas. *Environment and Planning B – Planning and Design, 35*, 480-503.

Gutiérrez, J., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, *19*(6), 1081-1092.

Horner, M. W., & Murray, A. T. (2004). Spatial representation and scale impacts in transit service assessment. *Environment and Planning B, 31*, 785-798.

Hsiao, S., Lu, J., Sterling, J., & Weatherford, M. (1997). Use of geographic information system for analysis of transit pedestrian access. *Transportation Research Record: Journal of the Transportation Research Board*, (1604), 50-59.

Jin, X., Beimborn, E., & Greenwald, M. (2004). *Impacts of accessibility, connectivity and mode captivity on transit choice* (No. Draft Report).

Keijer, M. J. N., & Rietveld, P. (2000). How do people get to the railway station? The Dutch experience. *Transportation Planning and Technology*, *23*(3), 215-235.

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice*, *38*(3), 223-247.

Liu, C., Erdogan, S., Ma, T., & Ducca, F. W. (2014). How to Increase Rail Ridership in Maryland? Direct Ridership Models (DRM) for Policy Guidance. In*Proceedings of 93rd Annual Meeting of the Transportation Research Board, Washington DC*.

Loutzenheiser, D. (1997). Pedestrian access to transit: model of walk trips and their design and urban form determinants around Bay area rapid transit stations. *Transportation Research Record: Journal of the Transportation Research Board*, (1604), 40-49.

Merkle, E. C., & Shaffer, V. A. (2011). Binary recursive partitioning: Background, methods, and application to psychology. *British Journal of Mathematical and Statistical Psychology*, *64*(1), 161-181.

Meyer, M. D., & Miller, E. J. (2001). *Urban transportation planning: a decision-oriented approach*.

Murray, A. T. (2001). Strategic analysis of public transport coverage. *Socio-Economic Planning Sciences*, *35*(3), 175-188.

Murray, A. T., Davis, R., Stimson, R. J., & Ferreira, L. (1998). Public transportation access. *Transportation Research Part D: Transport and Environment*, *3*(5), 319-328.

O'Neill, W. A., Ramsey, R. D., & Chou, J. (1992). Analysis of transit service areas using geographic information systems. *Transportation Research Record*, (1364).

O'Sullivan, S., & Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*, (1538), 19-26.

Parsons Brinckerhoff. (1996). Transit and Urban Form, (TCRP Report 16, vol. 1.) Transportation Research Board, National Research Council, Washington, DC.

Phillips, C.G., & Edwards, R. H. (2002). Socioeconomic, community-based approach for developing integrated mass transit systems: Application to City of Baltimore, Maryland. *Transportation Research Record: Journal of the Transportation Research Board*, (1793), 71-79.

Rhein, A. (2015, February 2). MARTA TOD Making Strides. *SaportaReport*, Retrieved from http://leadership.saportareport.com/

Taylor, B. D., & Fink, C. N. (2003). The factors influencing transit ridership: A review and analysis of the ridership literature. *University of California Transportation Center*.

Untermann, R. K. (1984). Accommodating the pedestrian: Adapting towns and neighbourhoods for walking and bicycling.

Walters, G., & Cervero, R. (2003). Forecasting transit demand in a fast growing corridor: the direct-ridership model approach. *Fehrs and Peers Associates*.

Zhao, F., Chow, L. F., Li, M. T., Ubaka, I., & Gan, A. (2003). Forecasting transit walk accessibility: regression model alternative to buffer method.*Transportation Research Record: Journal of the Transportation Research Board*, (1835), 34-41.

Zuehlke, K. (2007). Impossibility of Transit in Atlanta: GPS-Enabled Revealed-Drive Preferences and Modeled Transit Alternatives for Commute Atlanta Participants (Masters thesis). Retrieved from https://smartech.gatech.edu/handle/1853/19788